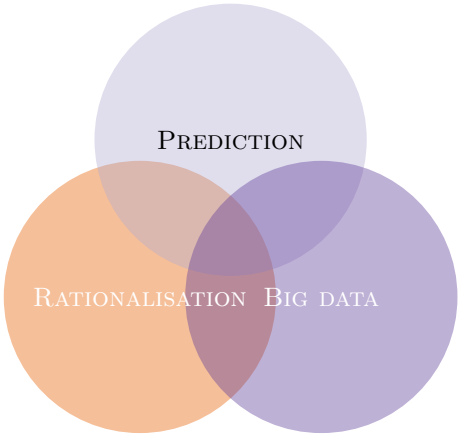
Understanding the Structure of Materials at the Intersection of Rationalisation, Prediction and Big Data

Martin Uhrin



University College London

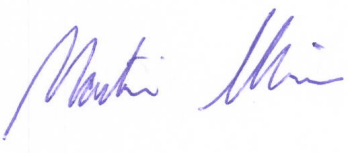
A dissertation submitted for the degree of

Doctor of Philosophy

August 2015

Declaration

I, Martin Uhrin, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. The work contains nothing which is the outcome of work done in collaboration except where specifically indicated in the text.



Martin Uhrin

August 2015

Abstract

Theoretical materials science has a large and growing role to play in modern society thanks to its ability to deliver materials with new and interesting properties. The properties of any material are, on some level, a function of its internal structure. In this work we combine three important tools spanning the last 100 years of materials research, rationalisation, prediction and big data in an attempt to understand the factors that underpin the stability of ordered structures and to build an understanding of structure that is agnostic of a particular element or building block. We apply rationalisation to data mining of the Inorganic Crystal Structure Database, using various proposed structure descriptors to probe the factors affecting structure stability. Extensive prediction is performed on the Fe-Ni-Si system at inner earth core pressures to determine the phases most likely to be present, yielding a new, stable, Ni-Si structure. A new prediction technique for 2D grain boundaries is presented that doubles the size of system that can reasonably be studied at the *ab initio* level of theory. The structurally rich phosphorus and arsenic systems are investigated using structure prediction, producing new metastable structures. Finally, we use a simple model for particles that attract at long range and repel at short to probe all the possible binary structures over a wide range of stoichiometries. By carrying out prediction over a wide range of potential parameters we build a database of almost 20M entries. Contained within are a number of unreported structures including many in parts of parameter space that go beyond the periodic table in terms of size and bond energy ratios. Our work provides hints that these hypothetical structures could be realised in self assembling systems made up from constituents with tunable interactions opening the door to the possibility of new properties.

Acknowledgements

No man is an Iland, intire of it selfe; every man is a peece of the Continent, a part of the maine...

John Donne, *Devotions upon Emergent Occasions*, 1624

Just as no (wo)man is an island, so no thesis stands alone. Anyone who has been in any way connected to the writing of a thesis knows this all too well. There is, of course, the connection to all the academic and scholarly activity that made this thesis possible, to the authors of which I am very grateful, but this is dealt with in the bibliography. Here, I take a moment to thank those to whom I have a personal connection, those who shared their time, energy and spirit with me during my time at UCL. These connections helped to shape me as a person, and will always be a part of me, so when you read these names know that they form a big part of the work you see in front of you.

My supervisor Chris, has been, and continues to be, an endless source of ideas and inspiration. Perhaps more importantly, he never lost faith in our direction and in me as a scientist (or if he did he never showed it!) which allowed us to take this work further than I ever thought possible and allowed me to reach a point where I could feel truly proud of my accomplishments. Andrew Morris, a name that always brings a smile to my face, was there from the start, sharing not only his intellect but his fantastic sense of humour. Miguel Martinez-Canales was always an inspiration with a remarkable breadth and depth of knowledge that he was always happy to share. Georg Schusteritsch, as excellent a scientist as he is a decent human being; it's hard to imagine how the last year and half would have been without him in the group.

My thoughts turn next to David Bowler, who was ever the safety net a good second supervisor should be. Allowing me to walk the PhD tightrope with confidence and composure until I would invariably fall, whereupon he was always ready and willing to help me back up.

Seto Balian's frequent visits to our office would always bring a smile to my face, even if it was just to do the recycling. Bobby Antonio, with his sharp, inquisitive and most of all open mind broadened my horizons immeasurably, not only in the world of science but also into the sphere of art.

I have made many great friends at UCL with whom I've shared some fantastic times. To mention a few more by name: Tassem El-Sayed, Clyde Fare, Steve Hepplestone, Andrew James, Kostas Konstantinou, Navaratnarajah Kuganathan, Costas Lazarou, Duncan Little, Shereif Mujahed, George Pender, Hugh Price, Kane Shenton and Cheah Wei Li.

Tania Montiero is someone I have great respect for. Always available for advice or just a chat and with a rare humbleness that melts away any distinction between student and professor.

I owe a great deal of thanks to Nicola Marzari who had the faith to hire me near the end of my PhD and give me the space I needed to finish this work.

My family have supported me throughput my journey. My brother, Dušan, shared many a weekend in London with me. He has a sense of adventure that's rare to find and a kindness of heart that's even rarer. My mother, Stanka, always gave me the freedom to grow and encouraged me to broaden my horizons. My father, also Dušan, was selflessly there whenever I needed him, often going far out of his way to ensure my well-being and happiness. My grandparents František and Tereza provided me with a sanctuary they called a "maternal ship", a home I could always come to and feel welcome and looked after. The thought of completing this journey without them is difficult to imagine.

Lastly I thank a very important person in my life, my companion and girlfriend, Cassie. We've shared many great years together including those at UCL. She's always encouraged me and been there during my times of doubt. I treasure the freeing feeling that comes with knowing that she is sharing her journey through life with me.

To all of those named above and many others: Thank you.

To Cassie

Contents

List of Figures

List of Tables

1 | Overview

The history of civilisation is often talked about, and indeed defined in terms of, materials ages. Having moved through the stone, bronze and iron ages we are currently in what can quite reasonably be described as the silicon age. The story of each age follows a similar pattern. The discovery of a new material precipitates a discontinuous change; suddenly we are able to do and make things that simply weren't possible before. This is followed by a period of mastery where knowledge is accumulated and skills honed enabling the new material to be applied more broadly and used in more intricate ways. Now, our ability to manipulate matter is so exquisite that we can move individual atoms and design electronic configurations that behave in exotic ways giving rise to new and sometimes counter-intuitive properties. The breadth of our ambition has also increased dramatically with modern materials covering almost the full range of the periodic table. Indeed, the average citizen of an industrialised nation will carry at least a third of the periodic table in their pocket alone. In this opinion of this author we are now entering an era decoupled from any particular material, where need and desire dictate properties which are in turn realised using the parts of the periodic table most suitable for the job. An era that some authors have called the "designed materials age".

This ability comes not a moment too soon. Today, civilisation is facing perhaps its biggest collective challenge with the threats posed by climate chaos on such a large and growing population. Undoubtedly many changes will need to happen at the social and political level but advanced materials certainly have a significant role to play. Areas such as power generation, distribution and storage, healthcare, desalination and transport to name a few have already benefited enormously thanks to the development of materials that exhibit new and useful properties, many of which are not found to occur naturally. To understand and predict the properties of any material it is essential to know the arrangement of atoms within it.

The central theme of this thesis is the understanding of the structures adopted by roughly spherical particles, be they atoms, interacting colloidal particles, polystyrene beads or others that have a propensity for forming regular assemblages. To do so we bring together three important tools.

Rationalisation

Patterns and sets of rules can be deduced using known structures which give an insight into the reasons for the formation of particular structures over others. If the rules are sufficiently robust they can even be used to make predictions about new structures.

Prediction

By using our knowledge of the interactions between structure forming constituents, we can use prediction to directly evaluate the relative stability of different structures to find the one most likely to form.

Big Data

Rather than stating what big data is, we define it in terms of what it enables one to do. That is, to pose questions over an enormous dataset of structures, the answers to which invariably raise more questions which in turn become new queries over the dataset. For this to be possible not only do queries need to be fast, but the database must be flexible enough to enable unforseen questions to be posed.

It may be tempting to view these as past, present and future because of their chronology, in reality all three bring a unique perspective that enhances our understanding of structure. In this work we bring the three together and break free from the constraints imposed by the periodic table, generating and analysing structures that have never been considered before.

This work is laid out as follows. After some necessary theory in chapter 2 we describe various attempts at structure rationalisation in chapter 3 and apply a multitude of proposed structure descriptors to current data from the Inorganic Crystal Structure Database. In chapter 4 we use structure prediction to determine the phases of the Fe-Ni-Si system at inner earth core pressures, an important step to understanding the formation of the earth. Chapter 5 looks at the structures formed by phosphorus and arsenic. These two systems are fascinatingly diverse in the structures they form, phosphorus in particular having more observed allotropes than almost any other element. In chapter 6 we present a new method for the prediction of 2D material interfaces that outperforms pure random structure searching by an order of magnitude, demonstrating the benefits of feeding back chemical knowledge to guide prediction. Finally, in chapter 7, we go beyond the periodic table building an encyclopedia of structures assuming only particles that attract at long range and repel upon coming together.

2 | Theory

2.1 Energy landscapes

The configuration space of a collection of N particles is spanned by a D dimensional vector, $\mathbf{x}$, where D is equal $3N-6$ (for $N>2$) or $(3N+1)$ for a periodic system. In the case of the latter six degrees of freedom define the unit cell, $3N$ the particle positions and three are redundant due to translational invariance. Given a potential energy function, $V(\mathbf{x})$, the potential energy surface (PES), also known as an energy landscape, is defined as the $D+1$ dimensional space formed by associating a potential energy with each point in configuration space.

A large part of any PES can be ignored in the context of predicting structure. Regions that contain disconnected monomers, molecules or clusters can be ignored when searching for crystal structures. Furthermore, a large portion of any PES is dominated by high energy structures that are much higher in energy than even the highest energy metastable structure. This is a consequence of the fact that many regions of the configuration space contain structures where two or more atoms are very close to each other. The large repulsive force between two such atoms gives rise to structures that are high in energy.

The remainder of the PES contains the structures of interest. This part of the energy landscape is dominated by basins of attraction containing metastable structures at their minima. A basin is defined by the set of points that will converge to a common minimum under the application of downhill relaxation. The stable structure at a given pressure is found at the bottom of the deepest basin. In several studies it has been found that deeper basins tend to have a larger hypervolume when compared to their higher energy counterparts (Doye and Massen, 2005; Doye, Wales and Miller, 1998). For structures with few atoms this can result in the lowest energy basin occupying a large portion of the energy landscape, a property that is exploited by some structure prediction techniques.

Another observed feature of energy landscapes is that neighbouring basins that have a large energy difference in their minima are more likely to have a small energy barrier. This is related to the Bell-Evans-Polanyi principle which states that the activation energy of highly exothermic reactions tends to be small. This also helps to explain the observed phenomenon that low energy

basins tend to be found together forming funnels in the energy landscape. For more complex structures there may be multiple disconnected funnels separated by high energy barriers.

Of crucial importance to structure prediction is the fact that the number of local minima scales exponentially with the number of atoms. This was observed in early computational studies on Lennard-Jones clusters by Hoare and McInnes (1976) and Tsai and Jordan (1993) and later explained by Stillinger (1999). The principle behind Stillinger's argument can summarised as follows. A sufficiently large system of N atoms can be divided into M identical subsystems. Each subsystem has an independent PES with a number of local minima given by $n(N/M)$. The total number of ways to combine the locally stable configurations is given by

$$n(N) = n^M(N/M), \tag{2.1}$$

which has the solution

$$n(N) = e^{\alpha N}, \tag{2.2}$$

where α is a unitless constant. For smaller systems that cannot necessarily be divided into identical subsystems, this relationship need not hold. Pickard and Needs (2011) report that when randomly sampling the PES of the Lennard-Jones solid the average number of structure relaxations needed to find the ground state grows linearly up to $N = 128$, after which exponential growth begins to dominate. Using a binary potential that models ionic interactions the same work found that exponential scaling takes hold above $N = 60$. These results can be understood to be a consequence of the competition between the respective growths of the hypervolume of the ground state basin and the total number of metastable basins.

2.1.1 Local optimisation

Local optimisation is an essential tool for exploring energy landscapes and forms an integral part of many global optimisation schemes. In general a local optimisation scheme takes as its input the current configuration of the system, $\mathbf{x}$, and returns a new state vector, $\mathbf{x}_{\min}$, such that $V(\mathbf{x}_{\min}) \leq V(\mathbf{x})$ and all the forces are zero i.e. the system is at a minimum. There are many such methods and here we will briefly mention some of those most commonly used for structure optimisation.

Nonlinear minimisation

In most cases, in addition to the energy for any point in configuration space it is possible to calculate forces acting on each particle, $\mathbf{g}(\mathbf{x}) = -\nabla V(\mathbf{x})$. In general forces are used in two ways during optimisation. The first is simply to know when a minimum has been reached, i.e. when $\|\mathbf{g}(\mathbf{x})\| < g_{\text{tol}}$. The second is to move the configuration closer to a minimum which in the general

can be written as

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha_n \mathbf{d}_n, \tag{2.3}$$

where n labels the current step, α_n is a stepsize and $\mathbf{d}_n$ is a search direction. The variation between optimisation algorithms comes almost entirely from the choice of α and $\mathbf{d}$.

Steepest descent

In the absence of additional information the simplest choice of search direction is simply $\mathbf{d} \equiv \mathbf{g}$, whereby the algorithm always moves in the direction of the force at the current point i.e. along the path of steepest descent. The stepsize, α, may either be a small fixed value, or calculated from a *line minimisation*. A line minimisation is a one-dimensional constrained optimisation used to identify the location of the minimum along the search direction. The resulting value of α will place the configuration at the bottom of the valley along the direction of $\mathbf{d}$ upon the application of eq. (2.3). While steepest descent is easy to implement it is seldom used as it tends to have poor convergence performance, converging linearly, and tends to "zigzag" along valleys. In certain pathological cases it may fail to reach the minimum all together.

Conjugate gradients

To improve upon steepest descent it is desirable to avoid search directions that have already been traversed, effectively limiting the amount of zigzagging the algorithm does as it homes in on the minimum. This is the basis of conjugate gradient methods. We begin by expanding our potential energy function around the desired minimum $\mathbf{h}$:

$$V(\mathbf{x}) = V(\mathbf{h}) + \frac{1}{2}(\mathbf{x}-\mathbf{h})^{\mathrm{T}}\mathbf{H}(\mathbf{x}-\mathbf{h}) + \text{higher terms}, \tag{2.4}$$

where $\mathbf{H}$ is the symmetric, positive definite, Hessian matrix, containing second-order partial derivatives of the energy. Taking the derivative and ignoring higher order terms the gradient is

$$\mathbf{g}(\mathbf{x}) = \mathbf{H}(\mathbf{x}-\mathbf{h}), \tag{2.5}$$

having used the symmetric property of $\mathbf{H}$. We now consider a series of n steps made according to eq. (2.3), for which we have

$$\mathbf{x}_n = \mathbf{x}_{j+1} + \sum_{i=j+1}^{n-1} \alpha_i \mathbf{d}_i, \tag{2.6}$$

for any $0 \leq j \leq n-1$. To simplify notation we let $\mathbf{g}(\mathbf{x}_i) = \mathbf{g}_i$. Pre-multiplying both sides of eq. (2.6) by $\mathbf{H}$, and using eq. (2.5), we get:

$$\mathbf{g}_n = \mathbf{g}_{j+1} + \sum_{i=j+1}^{n-1} \alpha_i \mathbf{H}\mathbf{d}_i. \tag{2.7}$$

To ensure orthogonality of the gradient at the new position and the previous search direction we require that

$$\mathbf{g}_{i+1}^{\mathrm{T}}\mathbf{d}_i = 0, \tag{2.8}$$

using which we get that

$$\mathbf{g}_n^{\mathrm{T}}\mathbf{d}_j = \sum_{i=j+1}^{n-1} \alpha_i \mathbf{d}_i^{\mathrm{T}}\mathbf{H}\mathbf{d}_j. \tag{2.9}$$

Now, let us assume that vectors $\mathbf{d}_0$, $\mathbf{d}_1$, $\mathbf{d}_2$, …, $\mathbf{d}_{n-1}$ are **H**-conjugate, satisfying

$$\mathbf{d}_i^{\mathrm{T}}\mathbf{H}\mathbf{d}_j = 0, \text{ for all } i \neq j, \tag{2.10}$$

then

$$\mathbf{g}_n^{\mathrm{T}}\mathbf{d}_j = 0 \tag{2.11}$$

and since $\mathbf{d}_0$, $\mathbf{d}_1$, $\mathbf{d}_2$, …, $\mathbf{d}_{n-1}$ form a basis

$$\mathbf{g}_n = 0 \tag{2.12}$$

and by eq. (2.5)

$$\mathbf{x}_n = \mathbf{h}. \tag{2.13}$$

Thus the minimum is reached on the n^{th} iteration (or earlier if any $\alpha_i = 0$) for any set of **H**-conjugate directions. In practice the conjugate directions can be calculated using quantities at hand, the current gradient and the previous direction,

$$\mathbf{d}_{n+1} = \mathbf{g}_{n+1} + \beta_{n+1}\mathbf{d}_n. \tag{2.14}$$

By defining $\mathbf{y}_n = \mathbf{H}\mathbf{d}_n$, a quantity available from line minimisation along $\mathbf{d}_n$, this can be rearranged to get

$$\beta_{n+1} = -\frac{\mathbf{y}_n^{\mathrm{T}}\mathbf{g}_{n+1}}{\mathbf{y}_n^{\mathrm{T}}\mathbf{d}_n}, \tag{2.15}$$

which, after some manipulation, can be shown to be

$$\beta_{n+1} = \frac{\mathbf{g}_{n+1}^{\mathrm{T}}\mathbf{g}_{n+1}}{\mathbf{g}_n^{\mathrm{T}}\mathbf{g}_n}. \tag{2.16}$$

For a quadratic functions, a reasonable approximation close to a minimum, conjugate gradients converge exactly quadratically to the minimum, saving a considerable number of iterations over steepest descent. Even for non-quadratic functions the savings are often considerable.

Quasi-Newton-Raphson

Steepest descent uses no information from previous steps when making moves. Conjugate gradients improves on this by taking into account previous directions, however, it conveniently sidesteps any explicit consideration of the curvature, as described by $\mathbf{H}$, during the descent. Newton-Raphson methods build on this by building up an approximation to the Hessian allowing more accurate conjugate directions to be chosen. Given a current approximation to the Hessian, $\mathbf{H}_i$, the search direction is given by

$$\mathbf{d}_i = -\mathbf{H}_i \mathbf{g}_i. \tag{2.17}$$

The Hessian itself it then updated at each step as

$$\mathbf{H}_{i+1} = \mathbf{H}_i + \frac{\mathbf{d}_i \mathbf{d}_i^{\mathrm{T}}}{\mathbf{d}_i^{\mathrm{T}} \mathbf{H}_i \mathbf{d}_i} - \frac{\mathbf{H}_i \Delta\mathbf{g}_i \Delta\mathbf{g}_i^{\mathrm{T}} \mathbf{H}_i}{\Delta\mathbf{g}_i^{\mathrm{T}} \mathbf{H}_i \Delta\mathbf{g}_i} \tag{2.18}$$

where

$$\Delta\mathbf{g}_i = \mathbf{g}_{i+1} - \mathbf{g}_i. \tag{2.19}$$

This method always converges and does so rapidly (Fletcher and Powell, 1963) however for large numbers of particles a considerable memory cost is incurred in storing $\mathbf{H}_i$ as this scales as N^2. A popular low memory variant called L-BFGS (Tuckerman, Berne and Martyna, 1992) is often employed which does not explicitly store the Hessian and has similar convergence characteristics.

Two-point steepest descent

The optimisation scheme used in many parts of this work is the Barzilai-Borwein two-point steepest descent method (Barzilai and Borwein, 1988) which combines aspects of steepest-descent and quasi-Newton methods. The key difference from steepest-descent is in the calculation of α_i which is given by minimising

$$\|\Delta\mathbf{x}_i - \alpha_i \Delta\mathbf{g}_i\|^2 \tag{2.20}$$

where

$$\Delta\mathbf{x}_i = \mathbf{x}_{i+1} - \mathbf{x}_i, \tag{2.21}$$

yielding

$$\alpha_i = \frac{\Delta\mathbf{x}_i^{\mathrm{T}} \Delta\mathbf{g}_i}{\Delta\mathbf{g}_i^{\mathrm{T}} \Delta\mathbf{g}_i}. \tag{2.22}$$

The choice of eq. (2.20) is made as it provides a two-point approximation to the secant equation that underlies quasi-Newton methods. By using an explicit form for α the need for a line minimisation is avoided making the method simple to implement. Testing by Barzilai and Borwein and others shows the method to have better convergence characteristics than steepest-descent while being

simultaneously more tolerant to ill-conditioning.

2.1.2 Global optimisation

At the heart of any structure prediction method lies a global optimisation algorithm that must search the PES with a view to finding the lowest energy structure. Inspiration for these has come from many sources and considerable progress has been made to tailor general optimisation algorithms to suit structure prediction.

Simulated annealing

Simulated annealing (SA), first described by Kirkpatrick, Gelatt and Vecchi (1983) and independently by Černý (1985), is an early example of a global optimisation method applied to structure prediction. SA takes inspiration from metallurgy where annealing involves heating a material followed by controlled cooling in an attempt to crystallise the atoms in a more ordered state and reduce the number of defects, thereby reducing the overall energy. As with many structure prediction methods one begins with a set of parameters describing the shape of the unit cell and a set of atomic coordinates (typically randomly generated). The energy surface is then sampled according the Metropolis Monte Carlo procedure (Metropolis et al., 1953) whereby a new configuration is generated near the current solution by giving the atoms a small random displacement. A move to this configuration is accepted with probability one if it is lower in energy than the old. Otherwise, the move is accepted with probability $\exp(-\Delta E/T)$, where ΔE is the energy difference between the configurations and T is the annealing temperature. A cooling schedule is imposed to bring the system from a high temperature, where almost all Metropolis moves are accepted, to a state where the system is frozen and lower energy configurations cannot be found. In the limit of infinitely slow cooling SA is guaranteed to find the global minimum (Laarhoven and Aarts, 1987), however with practical cooling rates the procedure often gets stuck in a local minimum. This is especially true if the global minimum is far away from the initial configuration and separated by high energy barriers. Despite this difficulty SA has been used successfully in the prediction of structures from experimental data (Pannetier et al., 1990) as well as in purely predictive studies (Schön and Jansen, 1996, 2001a,b).

Basin-hopping

Basin-hopping (Wales and Doye, 1997) involves a transformation of the PES into a set of catchment basins where a catchment is defined as the region of parameter space containing points that lead to a common minimum under the application of downhill relaxation. Crucially, the energy associated with each catchment is the energy of the minimum

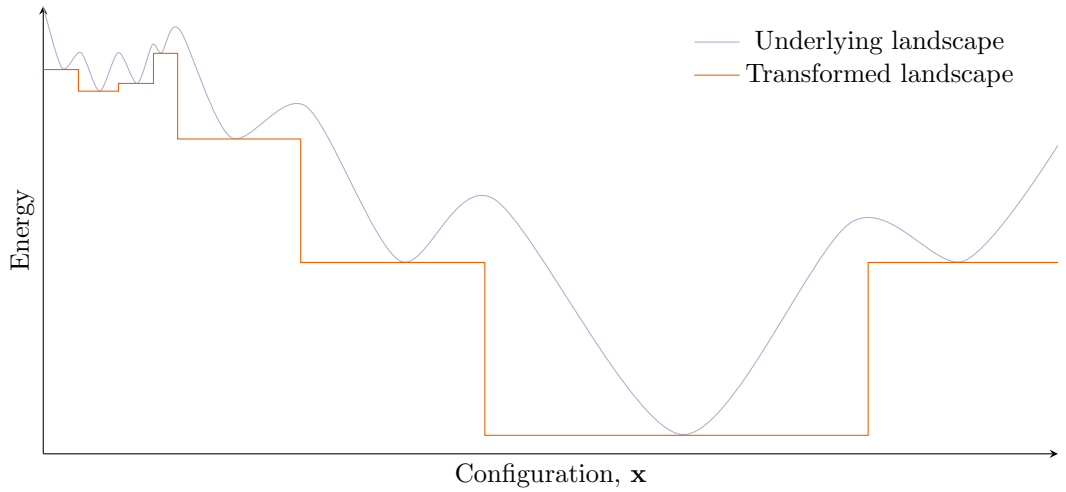


Figure 2.1: A schematic of the potential energy surface mapping applied in the basin hopping global optimisation scheme.

$$\tilde{V}(\mathbf{x}) = \min\{V(\mathbf{x})\}, \tag{2.23}$$

where min gives the energy after minimising from $\mathbf{x}$. This scheme partitions configuration space into a set of plateaux each with the energy of the associated minimum as shown in fig. 2.1. Unlike other methods such as metadynamics (Laio and Parrinello, 2002), basin hopping does not directly change the underlying landscape by altering the energy function, but rather acts as a mapping that is used during landscape exploration.

The most common search strategy employs Monte Carlo steps in the form of random perturbations made to the current configuration. Following such a step a local minimisation is performed and the step accepted if $V_{\text{new}} < V_{\text{old}}$ or if $\exp((V_{\text{old}} - V_{\text{new}})/kT)$ is greater than a random number drawn from the range [0,1). The second condition allows the algorithm to 'walk' out of funnels and explore different regions of the landscape.

The basin-hopping stepsize is greater than similar Monte Carlo simulations exploring thermodynamic properties as the goal is to move to nearby basins with each step. The maximum displacement for each coordinate is typically one third of the pair equilibrium distance (Wales, 2004), with the stepsize being dynamically adjusted to maintain a fixed acceptance ratio.

Evolutionary algorithms

Evolutionary algorithms (EAs) are based on principles borrowed from evolutionary biology, first applied to the problem of global optimisation in independent work by Friedman (1956), Box (1957), Blerdsoe (1961) and Bremermann (1962). The most commonly employed type of EA for structure prediction is the genetic algorithm (GA) where structures are represented as a binary string and have an associated fitness (the energy). Using this representation a diverse set of initial structures is generated, from which a subset is selected for *reproduction* with a bias towards the fittest. Pairs

of structures in this subset are then recombined, a process called crossover, by combining their structural features to produce child structures. Often this step is followed by *mutation*. Mutation typically involves a random perturbation of the atomic positions and can be vital in preventing *extinction* whereby the population diversity drops to such a degree that no new structures are being generated, effectively confining the search to one or more local minima. Finally, the population undergoes further selection and the whole procedure is repeated.

GAs have been successfully applied to a number of systems including metallic and alloy clusters, molecular crystals and metal oxides (Cartwright et al., 2004). The disadvantage of GAs is that they can vary greatly in their implementations and require a careful selection of crossover and mutation algorithms to produce the best outcomes for particular classes of system. Often it is not obvious how to tune these to prevent extinction, this makes GAs difficult to integrate as part of an automated procedure that can be used by non-domain experts.

An example of an implementation that tries to overcome some of these difficulties is the USPEX algorithm developed by Oganov, Ma et al. (2007). This is an example of an EA as structures are encoded as physical numbers as opposed to binary strings. The combined crossover and mutation step incorporates four possible operations: i) combining planar slabs from parent structures; ii) random atomic displacements; iii) swapping pairs of atoms of different species; and iv) special atomic displacements such as only moving low-order atoms. These have been shown to maintain good population diversity in binary Lennard-Jones clusters (Oganov, Lyakhov and Valle, 2011). Following crossover and mutation, duplicate structures are removed and the first n lowest energy structures are selected to be parents for the next generation. Oganov, Lyakhov and Valle (ibid.) reports that USPEX currently works well for structure of up to 30 atoms at the *ab initio* level of theory. USPEX has been used by Ma et al. (2009) to successfully predict a new insulating hP4 sodium structure that is stable above 250 GPa, an example of a fully ab initio prediction that was later confirmed by experiment.

Topological methods

Topological methods for structure prediction borrow ideas from graph theory to define infinite networks corresponding to hypothetical crystal structures. Instead of representing crystal structures as sets of points in space, this approach focuses on the network of chemical bonds between atoms. A graph is defined by a set of vertices and edges representing atoms and bonds respectively. A *net* is a specialisation of an infinite graph that is connected (there is a path between any two vertices) and simple (contains no loops, directed edges or multiple edges).

In early work Smith (1977, 1978, 1979) showed that this approach can be successfully used to predict framework structures such as zeolites. Chung, Hahn and Klee (1984) showed that by representing nets in a reduced form, as so called quotient graphs, it is possible to systematically generate nets. A quotient graph is finite and has vertices that correspond to the transnationally

invariant vertices of a net. The relationship between a quotient graph and a net is analogous to that between a primitive unit cell and the corresponding crystal structure. The power of this method is that specific crystal topologies can be systematically enumerated and tested for (meta)stability. Strong (2005), building on concepts developed by Thimm (2004), applied this method to the enumeration of *sp*-3 hybridised carbon polymorphs with up to four atoms per unit cell. As well as all of the known structures, three new structures were found that are potentially metastable.

Systematic enumeration of all possible quotient graphs is only possible for at most a handful of atoms in the unit cell as the combinatorial cost quickly becomes prohibitive. This limits the range of applicability of this approach when carrying out truly predictive studies. Further details on topological methods can be found in a nice introductory article by O'Keeffe (1999).

Random structure searching

The structure searching method used throughout this work is random structure searching (RSS), in contrast with the methods described thus far RSS is conceptually very simple. At its most basic, the procedure involves generating a unit cell with semi-random lattice parameters and populating it with atoms of the chosen species placed at random. The structure is then relaxed to find the bottom of the energy basin for that starting configuration. Optionally, the atomic coordinates may be perturbed and the structure relaxed again to try and find any nearby basins that are lower in energy. This is repeated using many different starting configurations in an attempt to sample diverse regions of the PES. Given the enormous size of the configuration space of a structure with just a handful of atoms it is somewhat surprising that such a simple approach works at all. However several of the properties of PES discussed in section 2.1 help to explain why it is successful. The fact that deeper basins tend to have a larger hypervolume greatly increases the chance of encountering the stable structure. Funnels group low energy basins, which can be exploited by perturbing relaxed structures. Furthermore, the fact that funnels can be separated by high energy regions favours an approach that samples the configuration space broadly.

Starting configurations Structure prediction, applied to clusters or bulk materials, is concerned with finding connected structures. It is therefore important to generate initial configurations where the atoms "feel" each other such that they join up during relaxation. Atomic diameters and bond lengths tend to fall in the range 0.75-3 Å, this is used as a guide for calculating the "natural volume" of an assembly of atoms. Random starting configurations are therefore generated using this volume, typically plus or minus 15%, to create so called "sensible starting structures".

Structure constraints Building on the basic algorithm one can add constraints to limit the types of initial configuration that are generated. Constraints such as minimum atoms separations, structure symmetries, minimum and maximum bond angles, coordination conditions and others

can be used to eliminate large portions of the search space and bias the search towards certain structures based on chemical intuition or experimental data. This can drastically reduce the time taken to find the ground state by effectively reducing the magnitude of the exponential scale factor, α, in eq. (2.2). In many cases these constraints are only imposed on the starting configurations leaving the structures unconstrained during the relaxation process. Nonetheless this simple step can significantly affect the distribution of structures found.

Termination criteria As with all structure searching methods applied to nontrivial structures, it is impossible to know for sure when the global minimum has been found, however a reasonable guess can often be made from experience gained when testing against similar structures of known ground state. In RSS the search is typically terminated when the current lowest energy structure has been found multiple times and no new structures can be found after several attempts.

Example applications The RSS algorithm is easy to implement and has very few tunable parameters that need to be chosen to suit particular classes of structure. Constraints are a powerful and physically meaningful way to favour certain structure types making it a much more intuitive method than those mentioned previously. It is also trivially parallelisable; Pickard and Needs (2011) report that on modern parallel computers *ab initio* random structure searching (AIRSS) works well for structures containing up to 12 atoms. Examples can be found in the literature of unconstrained searches with up the 48 atoms in the unit cell (Pickard and Needs, 2007). AIRSS combines random search with the plane-wave density functional theory code CASTEP to perform quantum mechanical structure relaxations to maximise its predictive potential. The versatile nature of AIRSS has seen it applied to a range of structure types.

Pickard and Needs (ibid.) used AIRSS to study the phase diagram of hydrogen and found several candidate structures for phase II based on molecular packing on distorted Hexagonal Close-Packed (*hcp*) lattices. For phase III they predict a new insulating molecular phase stable at pressures well above 300 GPa that is compatible with available experimental data.

In an early application of symmetry constraints Pickard and Needs (2008) studied high-pressure crystalline phases of ammonia. To generate initial configurations, chemical units were placed randomly within the unit cell subject to a set of symmetry operations. They found that between 90 and 440 GPa ammonia tends to form ionic structures consisting of NH_4^+ and NH_2^- ions as opposed to the typical hydrogen-bonded phases.

Srepusharawoot et al. (2010) applied AIRSS to study alkaline earth dicarbide structures. Structures were generated by imposing a set of symmetry operations from a randomly chosen space group. This is reported to have resulted in an estimated one order of magnitude gain in efficiency due to the reduction in the size of Brillouin zones. They found the lowest energy configuration for BeC_2 and MgC_2 to be a chain composed of C_5 rings connected by a single carbon atoms with Be or

Mg ions trapped between the chains. These are both previously unknown ground state structures. The stable configuration of the remaining XC_2, X = Ca-Ba, compositions was found to be the experimentally known chain structure composed of repeating X-C=C units.

More recently Mulroue, Morris and Duffy (2011) used an extension of AIRSS developed by Morris, Pickard and Needs (2008) to find the site of the stable interstitial defect in charged-state zirconolite. Zirconolite has possible uses in the disposal of highly active nuclear waste and it is therefore important to understand how it is effected by radiation damage. To find the stable defect site they introduced a randomly placed interstitial atom of the desired charge state into an 88 atom unit cell and relaxed the structure. This was repeated using 100 different starting configurations to identify the structure with the lowest energy. Once this was found a supercell of twice the size was created to be able to calculate the Frenkel defect energies.

2.1.3 Determining composition stability

In general, to determine the stability of a particular structure it is not enough to know that it is ground state of a particular landscape. It may be unstable to demixing into two or more competing phases of different stoichiometry, forming an interface in-between. The formation energy for a mixed phase with total energy E is given by

$$E_{\mathrm{f}} = E - \sum_{\alpha} n_{\alpha}\mu_{\alpha}, \tag{2.24}$$

where n_α are the numbers of particles of species α which have chemical potentials μ_α. The chemical potentials are obtained from the ground state of pure α as $\mu_\alpha = E_\alpha/n_\alpha$. For bulk systems the choice of these "endpoints" is usually obvious, however for molecular systems it often makes more sense to choose the molecular crystal such as in the case of oxygen. The stability of different phases can then be compared by calculating the formation energy per atom, given by

$$E_{\mathrm{pa}} = \frac{E - \sum_\alpha n_\alpha \mu_\alpha}{\sum_\alpha n_\alpha}. \tag{2.25}$$

A convex hull diagram, an example of which is shown in fig. 2.2, can be used to visualise the stable compositions. The stable structures are found on the hull with "tie-lines" joining adjacent stable points. Points that lie above the hull have a formation energy that favours demixing into the two (or more) phases surrounding the intersection point when projecting down onto the hull. Whether this will happen or not in practice depends the interface formation energy.

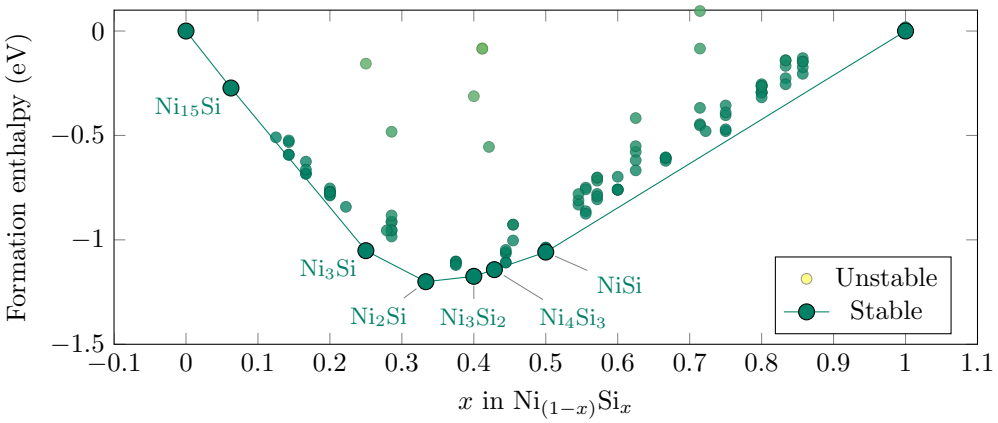


Figure 2.2: An example of a convex hull showing the stable phases of the Ni-Si system at 350 GPa.

2.2 Electronic structure

Up until now we have largely ignored the choice of $V(\mathbf{x})$ that defines the energy and forces through the interactions between the particles in our system. Empirical potentials, using a fixed functional form for the interaction between atoms, are widely used in materials modelling and even structure prediction. Indeed, in a later chapter we use the Lennard-Jones potential to explore crystal structures in a way that is agnostic of particular atomic species. However, to be able explore new configurations and predict truly new structures of atomic systems there are few substitutes for a quantum mechanical treatment. Ultimately this adaptability comes from an explicit treatment of the electrons which can respond to their chemical environment and dictate the bonding that in turn drives the configuration adopted by the ions in their ground state. In this section we cover some important aspects of electronic structure, in particular density functional theory which is used in several chapters.

2.2.1 Constructing the Hamiltonian

The general many-body Schrödinger equation for a solid can be written as

$$\hat{H}\Psi(\{\mathbf{R}_I;\mathbf{r}_i\}) = E\Psi(\{\mathbf{R}_I;\mathbf{r}_i\}), \tag{2.26}$$

where $\mathbf{R}_I$ and $\mathbf{r}_i$ are vectors that represent the ionic and electronic positions respectively. The Hamiltonian can then be constructed in the usual way, as the sum of the kinetic and potential energy terms. The kinetic energy operator is simply

$$\hat{T} = -\sum_I \frac{\hbar^2}{2M_I}\nabla^2_{\mathbf{R}_I} - \sum_i \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i},$$

while the potential energy operator is

$$\hat{V} = \frac{1}{2}\sum_{i,j(i\neq j)} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|} + \frac{1}{2}\sum_{I,J(J\neq I)} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}$$

where Z_I are the valence charges and e is the electronic charge. These three terms represent the energy associated with the repulsion between pairs of electrons, attraction between ion-electron pairs and repulsion between pairs of ions respectively. At this stage we apply the Born-Oppenheimer approximation which assumes that the electrons adapt instantaneously and adiabatically in response to any motion of the ions. As far as the quantum mechanics is concerned, we therefore assume that the ions are at rest and treat any ionic motion classically leaving the Hamiltonian as

$$\hat{H}^{\mathrm{BO}} = -\sum_i \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} + \frac{1}{2}\sum_{i,j(i\neq j)} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|} + \frac{1}{2}\sum_{I,J(J\neq I)} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}. \tag{2.27}$$

Furthermore, the last term may be neglected for the moment as it is simply an energy constant as far as the electronic degrees of freedom are concerned. This Hamiltonian remains difficult to solve for all but the most trivial of systems. In particular there are two main difficulties that arise. The first is that electrons, as spin-$\frac{1}{2}$ particles, have anti-symmetric wavefunctions and therefore Ψ must change sign if any two electrons interchange positions. Physically this is a manifestation of the Pauli exclusion principle and is known as the *exchange* problem. The second is simply the many-body nature of the system whereby the motion of every electron is affected by the motion of every other electron, this is known as the *correlation* problem.

2.2.2 The Hartree approximation

The Hartree approximation deals with the problem of correlation by splitting up the wavefunction into a product of non-interacting wavefunctions, thus

$$\Psi^{\mathrm{H}}(\{\mathbf{r}_i\}) = \phi(\mathbf{r}_1)\phi(\mathbf{r}_2)\ldots\phi(\mathbf{r}_N). \tag{2.28}$$

The total energy is now

$$\begin{aligned} E^{\mathrm{H}} &= \langle\Psi^{\mathrm{H}}|\,\hat{H}^{\mathrm{BO}}\,|\Psi^{\mathrm{H}}\rangle \\ &= \sum_i \left\langle \phi_i \left| \frac{-\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} - \sum_I \frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|} \right| \phi_i \right\rangle + \frac{e^2}{2}\sum_{i,j(i\neq j)} \left\langle \phi_i\phi_j \left| \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right| \phi_i\phi_j \right\rangle. \end{aligned}$$

From this we can get the so called single-particle Hartree equations by using a variational argument which come out as

$$\left[\frac{-\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} - \sum_I \frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|} + e^2\sum_{j(j\neq i)} \left\langle \phi_j \left| \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right| \phi_j \right\rangle\right] \phi_i(\mathbf{r}_i) = \epsilon_i\phi_i(\mathbf{r}_i), \tag{2.29}$$

where ϵ_i are Lagrange multipliers which act as normalisers of the single particle states. It is now possible to solve for each orbital $\phi_i(\mathbf{r}_i)$ separately, *given* all of the other orbitals $\phi_j(\mathbf{r}_j)$, $j \neq i$. Of course these in turn depend on ϕ_i, however the entire set of orbitals can be solved for by adopting an iterative self-consistent approach whereby an initial guess for the set of ϕ_ks is made which is then used to solve for each ϕ_i separately. These are in turn fed back and the procedure repeated until the variation in ϕ_ks drops below some tolerance.

2.2.3 The Hartree-Fock approximation

Returning to the problem of exchange mentioned earlier it is possible to use Slater determinants (Slater, 1929) to guarantee that our wavefunction meets the required symmetry conditions. This new wavefunction is written as

$$\Psi^{\mathrm{HF}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_1(\mathbf{r}_2) & \cdots & \phi_1(\mathbf{r}_N) \\ \phi_2(\mathbf{r}_1) & \phi_2(\mathbf{r}_2) & \cdots & \phi_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1) & \phi_N(\mathbf{r}_2) & \dots & \phi_N(\mathbf{r}_N) \end{vmatrix}, \tag{2.30}$$

and the corresponding total energy is

$$\begin{aligned} E^{\mathrm{H}} &= \langle \Psi^{\mathrm{HF}} | \, \hat{H}^{\mathrm{BO}} \, | \Psi^{\mathrm{HF}} \rangle \\ &= \sum_i \left\langle \phi_i \left| \sum_i \frac{-\hbar^2}{2m_e} \nabla^2_{\mathbf{r}_i} - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|} \right| \phi_i \right\rangle \\ &\quad + \frac{e^2}{2} \sum_{i,j(i\neq j)} \left\langle \phi_i \phi_j \left| \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right| \phi_i \phi_j \right\rangle \\ &\quad - \frac{e^2}{2} \sum_{i,j(i\neq j)} \left\langle \phi_i \phi_j \left| \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right| \phi_j \phi_i \right\rangle . \end{aligned} \tag{2.31}$$

Solving as we did before, the single-particle Hartree-Fock (HF) equations are

$$\begin{aligned} \left[\frac{-\hbar^2}{2m_e} \nabla^2_{\mathbf{r}_i} - \sum_I \frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|} + e^2 \sum_{j(j\neq i)} \left\langle \phi_j \left| \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right| \phi_j \right\rangle \right] \phi_i(\mathbf{r}_i) \\ - e^2 \sum_{j(j\neq i)} \left\langle \phi_j \left| \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right| \phi_i \right\rangle \phi_j(\mathbf{r}_i) = \epsilon_i \phi_i(\mathbf{r}_i). \end{aligned} \tag{2.32}$$

Compared to eq. (2.29) the final term on the left hand site is new, this so called "exchange" term accounts for the Fermi repulsion between pairs of electrons. It should be noted that in this analysis we have neglected spin for the sake of simplicity, however were it included the result would be much the same with an explicit spin degree of freedom in addition to the spatial.

2.2.4 Density functional theory

In the previous analysis we made a rather drastic approximation in eq. (2.28). By splitting up the wavefunction in this way each electron is assumed to be non-interacting and is affected by the other electrons in the system only through the mean field. Density functional theory (DFT), formulated in two seminal papers by Hohenberg, Kohn and Sham (Hohenberg and Kohn, 1964; Kohn and Sham, 1965), takes a different approach deriving exact single particle wavefunctions with approximations introduced as needed. The key step is to abandon the representation of the many-body wavefunction, $\Psi(\mathbf{r}_i)$, as a function of the explicit position of each electron and instead to reformulate the problem in terms of the total electron density, $n(\mathbf{r})$. In doing so, as we shall show, it is possible to avoid explicitly specifying the full many-body wavefunction.

Hohenberg-Kohn

Hohenberg and Kohn begin by defining an exact theory for many-body systems of interacting particles in an external potential, $V(\mathbf{x})$, in this case applied to electrons and fixed nuclei. To be able to formulate the theory in terms of the ground state density, $n(\mathbf{r})$, it is first necessary to show that this is uniquely define (up to a constant) by the external potential. To proceed let us assume the opposite and suppose that there exist two different external potentials, $V(\mathbf{r})$ and $V'(\mathbf{r})$, that give rise to the same ground state electron density, $n(\mathbf{r}) = n'(\mathbf{r})$. The energies for these two systems can be written as

$$E = \langle\Psi|\hat{H}|\Psi\rangle \text{ and } E' = \langle\Psi'|\hat{H}'|\Psi'\rangle\,, \tag{2.33}$$

where the primed symbols refer to the system with potential $V'(\mathbf{r})$. Given that Ψ' is not the ground state of H, we know that

$$\langle\Psi|\hat{H}|\Psi\rangle < \langle\Psi'|\hat{H}|\Psi'\rangle\,, \tag{2.34}$$

where the strict inequality holds assuming a non-degenerate ground state. Using the variational principle it follows that

$$\begin{aligned} E < \langle\Psi'|\hat{H}|\Psi'\rangle &= \langle\Psi'|\hat{H} + \hat{V}' - \hat{V}'|\Psi'\rangle = \langle\Psi'|\hat{H}' + \hat{V} - \hat{V}'|\Psi'\rangle \\ &= \langle\Psi'|\hat{H}'|\Psi'\rangle + \langle\Psi'|(\hat{V} - \hat{V}')|\Psi'\rangle \\ &= E' + \langle\Psi'|(\hat{V} - \hat{V}')|\Psi'\rangle\,. \end{aligned} \tag{2.35}$$

Equally we can show that

$$E' < E - \langle\Psi|(\hat{V} - \hat{V}')|\Psi\rangle\,. \tag{2.36}$$

Adding together eqs. (2.35) and (2.36) we get

$$E + E' < E + E' + \langle\Psi'|(\hat{V} - \hat{V}')|\Psi'\rangle + \langle\Psi|(\hat{V} - \hat{V}')|\Psi\rangle\,, \tag{2.37}$$

for which the last two terms are

$$\int n'(\mathbf{r})(V(\mathbf{r}) - V'(\mathbf{r}))\mathrm{d}\mathbf{r} - \int n(\mathbf{r})(V(\mathbf{r}) - V'(\mathbf{r}))\mathrm{d}\mathbf{r}, \tag{2.38}$$

but by our original assumption $n(\mathbf{r}) = n'(\mathbf{r})$ and therefore this term is zero leaving the inequality $E + E' < E + E'$. From this we conclude that the original assumption must be wrong and that no two (nontrivially related) external potentials can give rise to the same ground state density, $n(\mathbf{r})$. Equally, this means that the ground state density defines the external potential (up to a constant) and in turn the Hamiltonian.

Once the Hamiltonian is known it is, in principle, possible to determine the wavefunction for any state by solving the Schrödinger equation. Thus we note the remarkable fact that the ground state density determines all the ground *and* excited state properties of the full many-body system.

Next we show that a universal functional for the energy, valid for any external potential, can be defined in terms of $n(\mathbf{r})$. Here we follow the argument put forward by Levy (1979) and Lieb (1983). The total energy for any many-body wavefunction can be written as

$$E = \langle\Psi|\hat{T}|\Psi\rangle + \langle\Psi|\hat{V}_{\mathrm{int}}|\Psi\rangle + \int V(\mathbf{r})n(\mathbf{r})\mathrm{d}\mathbf{r}, \tag{2.39}$$

where V_{int} represents the electron-electron interactions. Next we consider the set of many-body wavefunctions that correspond to a particular density, $n(\mathbf{r})$, the unique lowest energy for which can be found by performing the minimisation

$$E_{\mathrm{LL}}[n] = \min_{\Psi\to n(\mathbf{r})} [\langle\Psi|\hat{T}|\Psi\rangle + \langle\Psi|\hat{V}_{\mathrm{int}}|\Psi\rangle] + \int V(\mathbf{r})n(\mathbf{r})\mathrm{d}\mathbf{r}, \tag{2.40}$$

from which the Levy-Lieb functional is defined as

$$F_{\mathrm{LL}}[n] = \min_{\Psi\to n(\mathbf{r})} [\langle\Psi|\hat{T} + \hat{V}_{\mathrm{int}}|\Psi\rangle] \tag{2.41}$$

This functional is universal in the sense that is common to all solids as it is independent of the external potential.

Kohn-Sham

The problem with eq. (2.41) is that we have no way to know how to arrive at the universal functional, only that it exists. To be able to carry out calculations we turn to the Kohn-Sham approach. Kohn and Sham make the ansatz that there exists a non-interacting system whose ground state density is the same as that of the full many-body system. We define the classical

Coulomb energy of the electron density (the Hartree energy) as

$$E^{\mathrm{H}}[n] = \frac{1}{2}\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'. \tag{2.42}$$

Using the kinetic energy functional of the non-interacting system, $T^{\mathrm{S}}[n]$, the ground state energy functional can now be written as

$$\begin{aligned} E^{\mathrm{KS}} &= \langle\hat{T}\rangle + \langle\hat{V}_{\mathrm{int}}\rangle + \int V(\mathbf{r})n(\mathbf{r})\mathrm{d}\mathbf{r} \\ &= T^{\mathrm{S}}[n] + \int V(\mathbf{r})n(\mathbf{r})\mathrm{d}\mathbf{r} + E^{\mathrm{H}}[n] + E^{\mathrm{XC}}[n] \end{aligned} \tag{2.43}$$

where the exchange-correlation (XC) functional can be written as

$$E^{\mathrm{XC}}[n] = \langle\hat{T}\rangle - T^{\mathrm{S}}[n] + \langle\hat{V}_{\mathrm{int}}\rangle - E^{\mathrm{H}}[n]. \tag{2.44}$$

In the above form it is clear that this accounts for the differences in kinetic and potential energy due to the unknown many-body exchange-correlation effects, which are in general relatively small compared to the total energy. The independent-particle kinetic energy, T^{S}, is however not small and the form of the functional is not known. Therefore, we must resort to solving the single particle equations given by

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}} + V^{\mathrm{eff}}(\mathbf{r}, n(\mathbf{r}))\right]\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}), \tag{2.45}$$

where the effective potential, V^{eff}, is defined as

$$V^{\mathrm{eff}}(\mathbf{r}, n(\mathbf{r})) = V(\mathbf{r}) + e^2\int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}' + \frac{\delta E^{\mathrm{XC}}[n(\mathbf{r})]}{\delta n(\mathbf{r})}. \tag{2.46}$$

The set of eq. (2.45) are called the Kohn-Sham equations and their solutions the Kohn-Sham orbitals. Again, as in Hartree-Fock, the these equations must solved for iteratively until self-consistency is reached. These can then be used to get the corresponding kinetic energy as

$$T^{\mathrm{S}} = -\frac{\hbar^2}{2m_e}\sum_i \langle\phi_i|\nabla^2|\phi_i\rangle \tag{2.47}$$

Because the orbitals and energy eigenvalues, ϕ_i and ϵ_i, correspond to a set of fictitious particles they have no physical meaning with the exception of the highest eigenvalue in a finite system. This corresponds to the negative of the ionisation energy. In theory this means that calculated band gaps should also have little physical meaning, however empirically it has been found that the values can be reasonable, typically being underestimated.

Exchange-correlation functionals

One crucial piece which remains unaccounted for is E^{XC}. Unfortunately the Kohn-Sham equations give no indication of how to derive it, we must therefore make an educated guess. Many suggestions have been made, perhaps the simplest of which is the local-density approximation (LDA). Under this approximation the exchange-correlation energy density at every point is assumed to be the same as that for a homogeneous electron gas with the same density, i.e.

$$E^{\mathrm{XC}}_{\mathrm{LDA}} = \int n(\mathbf{r})\epsilon^{\mathrm{XC}}_{\mathrm{hom}}(n(\mathbf{r}))\mathrm{d}\mathbf{r}. \tag{2.48}$$

This approximation works best for metals, where the nearly-free-electrons are reasonably approximated as a homogeneous gas, and worst for strongly covalent solids with an inhomogeneous electron density caused by directional bonding.

The LDA uses only the density at every point in space, we may consider incorporating the gradient in an attempt to improve accuracy, particularly for systems that are poorly approximated as a homogeneous electron gas. This is the so called generalised-gradient approximation (GGA) where the energy functional is given by

$$E^{\mathrm{XC}}_{\mathrm{GGA}} = \int n(\mathbf{r})\epsilon^{\mathrm{XC}}_{\mathrm{hom}}(n(\mathbf{r}), |\nabla n(\mathbf{r})|)\mathrm{d}\mathbf{r}. \tag{2.49}$$

For many systems this approximation gives a marked improvement in accuracy over LDA, however it does not represent a universal improvement leading to worse result in certain cases, particularly in systems where the gradient changes very quickly.

A host of other functionals have been proposed, often to deal with particular shortcomings such as the underestimation of bandgaps or to explicitly take into account particular quantum effects, a good comparison of some of these can be found in R. Martin (2008). The problem with the profusion of functionals is that often there is not enough understanding of their accuracy across a range of systems which leads to a temptation to use the functional that gives the desired answer. Both LDA and GGA are simple with well understood behaviour, for this reason they are often a good choice to give a baseline results that can be easily compared.

Dispersion corrections

Exchange-correlation functionals try to account for the effects of many-body correlations between fragments of electron density in the system. However, they only treat static correlations, thus even if the exact exchange functional were known, dynamical correlations would remain unaccounted for. This poses a problem for systems with dispersion interactions, often referred to as van der Waals interaction, which arise because of fluctuations in one fragment of electron density that causes a coupled response in another via the generated electric field. The leading term of this attractive

dipole-dipole interaction is the familiar $-1/r^6$ decay found in, for example, the Lennard-Jones potential. A number of extensions to DFT, so called dispersion corrections, have been proposed to account for this missing interaction. Perhaps the simplest class, often termed "DFT-D", simply add a dispersion correction energy to the DFT energy to get a new total energy, $E_{\text{tot}} = E_{\text{DFT}} + E_{\text{disp}}$, where

$$E_{\text{disp}} = -\sum_{\text{A,B}} C_6^{\text{AB}} / r_{\text{AB}}^6, \tag{2.50}$$

and C_6^{AB} are the dispersion coefficients between species pairs A and B. These are often determined from various formulae involving experimental quantities. An example of a popular scheme based on this approach is called DFT-D2, developed by Grimme (2006). In chapter 5 we use a similar scheme called DFT-TS (Tkatchenko and Scheffler, 2009) which goes somewhat beyond the previously outlined approach by allowing the dispersion coefficients to vary the with the atomic environment. In DFT-TS the electron density is partitioned between each atom and compared to the density of the corresponding free atom. The difference is used to scale the C_6 coefficients allowing the interaction strength to vary in response to, for example, changes in the bonding environment. The accuracy of such approaches tends to be appreciably better than those that use fixed C_6 coefficients while incurring only a modest computational cost, depending on the partitioning scheme used. There are generally more accurate schemes that can be employed that try to take into account long-range density fluctuations and beyond pairwise interactions, however these tend to come with a much higher computational cost. A review by Klimeš and Michaelides (2012) gives a more complete overview including a comparison of relative errors.

Pseudopotentials

Pseudopotentials represent an important practical tool that can be used to significantly speed up electronic structure calculations with almost no impact on the overall accuracy if used correctly. The basic idea is to treat core and valence electrons separately. This choice is motivated by the fact that valence electrons, suitably defined, are responsible for bonding while the core electrons act primarily to screen them from the nuclear charge. A pseudopotential replaces the effects of the nucleus and core electrons up to a cutoff radius beyond which it is matched to the real wavefunction.

The modern pseudopotential approach was established by Phillips and Kleinman (1959) who proposed a nodeless wavefunction, $\tilde{\phi}_v$, given by

$$|\tilde{\phi}_v\rangle = |\phi_v\rangle + \sum_c \alpha_{cv} |\phi_c\rangle , \tag{2.51}$$

where $\alpha_{cv} = \langle \phi_c | \tilde{\phi}_v \rangle \neq 0$ and subscripts c and v respectively label the chosen core and valence states of the single particle wavefunctions, ϕ_i. By acting on eq. (2.51) with $\hat{H}$ and rearranging we

get the modified Schrödinger equation

$$\left[\hat{H} + \sum_c (\epsilon_v - \epsilon_c) \left|\phi_c\right\rangle \left\langle\phi_c\right|\right] \left|\tilde{\phi_v}\right\rangle = \epsilon_v \left|\tilde{\phi_v}\right\rangle \tag{2.52}$$

which demonstrates that it is possible to generate a pseudo-Hamiltonian, $\hat{H}^{\mathrm{PS}} = H + \sum_c (\epsilon_v - \epsilon_c) \left|\phi_c\right\rangle \left\langle\phi_c\right|$, with the same eigenvalues as the original but with the benefits of a smoother, nodeless wavefunction. Using the bare nuclear potential, $\hat{V} = Z_I e / r\hat{I}$, we get the associated pseudopotential,

$$\hat{V}^{\mathrm{PS}} = \frac{Z_I e}{r}\hat{I} + \sum_c (\epsilon_v - \epsilon_c) \left|\phi_c\right\rangle \left\langle\phi_c\right| . \tag{2.53}$$

Given that a pseudopotential must describe not only the nucleus but also the core electrons it cannot be simply a function a distance. In general the complexity of the core electrons can be captured by using a so called semi-local potential where a different potential is used for each angular momentum value given by

$$\hat{V}^{\mathrm{SL}} = \sum_{lm} \left|Y_{lm}\right\rangle V_l(r) \left\langle Y_{lm}\right| , \tag{2.54}$$

where $Y_{lm}(\theta, \phi) = P_l(\cos(\theta)) \exp(\mathrm{i}m\phi)$.

A pseudopotential is defined up to a particular core radius, r_c, after which it exactly matches the nuclear potential. Generally the smaller the core radius the better the potential will match the nuclear potential, however this comes with the associated cost of requiring a larger basis to describe the region outside r_c. A class of pseudopotential called norm-conserving require that the integral of the pseudo-wavefunctions up to r_c match that of the all-electron wavefunctions, equalling the total number of electrons in that domain. This development greatly improved the accuracy and transferability of pseudopotentials (Brazdova and Bowler, 2013) at the cost of generally leading to harder pseudopotentials that vary more rapidly and are more computationally taxing. The development of so called ultrasoft pseudopotentials by Vanderbilt (1990) and the closely related projector augmented wave method by Blöchl (1990) permitted the removal of the norm conservation requirement by introducing compensation charges to restore the norm. These provide considerable efficiency gain at the cost of being more complex to implement.

3 | Understanding structure

3.1 Introduction

The discovery of x-ray diffraction in crystals by Max von Laue (Friedrich, Knipping and Laue, 1913) saw an explosion in the number of crystal structures known to science. Prior to this most of our knowledge was based on symmetry arguments applied to the stacking of spheres, the most notable example being the work of Barlow (1883) who proposed 5 ionic structures that were later confirmed by x-ray. Today the Inorganic Crystal Structure Database is replete with over 170,000 structures almost entirely derived from x-ray diffraction. In common with many scientific discoveries von Laue's work lead to many more questions than answers. With an abundance of structure data it was not long before many great minds put themselves to work deducing and deciphering the patterns within. The purpose of this chapter is to highlight some of the progress that has been made and to emphasise ideas and concepts that are particularly relevant to the work of chapter chapter 7 which follows a similar route of discovery but using data generated from high-throughput computing as opposed to x-ray crystallography.

A great deal of progress has been made since Maddox (1988) controversially wrote of the 'scandal' that even simple crystalline solids could not be predicted given the composition alone. So much so that prediction of crystal structure by solving tractable yet accurate approximations to the Schrödinger equation is now a common research activity. Yet the outputs of such calculations rarely betray the reasons for the relative stability of competing phases. Efforts to find tangible factors such as 'atom size' or 'interaction strength' to give some semi-quantitative understanding have a long history and continue to this day. The work of Goldschmidt (1929) represents the first comprehensive undertaking of this nature. Goldschmidt grouped structures by a) the stoichiometry, b) the atomic size ratios, and c) their polarisations. Amongst his findings he showed that the structure of difluorides and dioxides can be well separated by considering the ionic size ratios alone as shown in fig. 3.1. Stoichiometry and some measure of atomic size remain common quantities in use today, while polarisation is more commonly replaced with quantities derived from electronic structure.

Further progress was made by Pauling (1960) who discussed rules governing the packing of

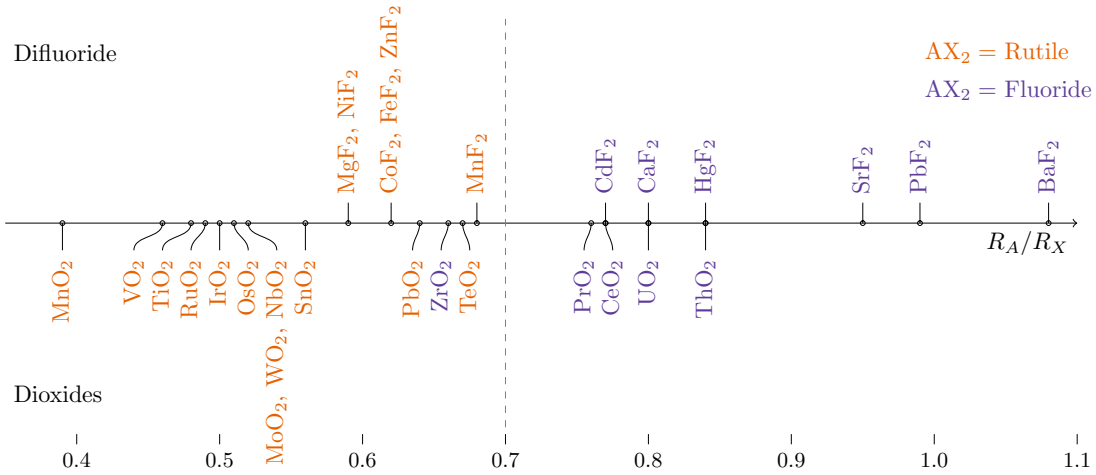
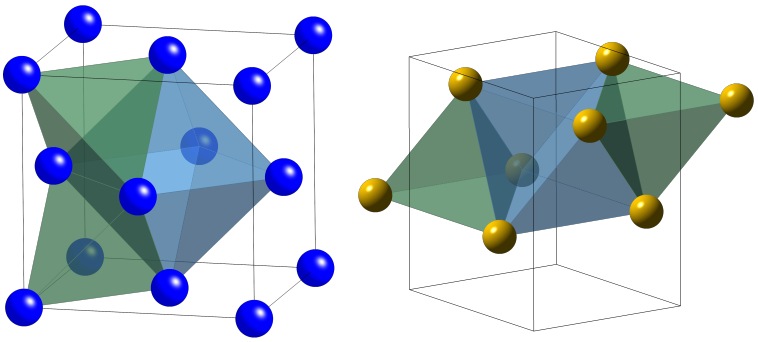


Figure 3.1: Structures of difluoride and dioxides against ionic radius ratios as determined by Goldschmidt (1929). Around a ratio of 0.7 structures in both groups are found to change from rutile to fluorite.

anions and cations in ionic crystals in terms of charges, radii, coordination polyhedra and stoichiometries. Around the same time metallic crystals were rationalised by Hume-Rothery (1955) by considering ratios of atomic size, electronegativity and number of valence electron. These concepts were brought together by Laves (1955) who attempted to synthesise one set of rules to govern all solids, metallic and non-metallic alike. In this seminal work he set out three principles governing structures: a) that space be filled as efficiently as possible, b) that packings tend to adopt high symmetry, and c) that connections between constituents tend to maximise connectivity. An important example of the application of these principles is the work Frank and Kasper (1958) which is discussed in detail in section 3.2.

Developments that followed started to move away from the largely qualitative descriptions that preceded and onto more quantitative ground. In particular structure maps, where structures types are plotted as a function of two or more parameters (e.g. electronegativity difference, mean quantum number, mean bandgap, etc), became increasingly common. Some of the earliest examples are provided by Mooser and Pearson (1959) and by Phillips (1970) who plotted maps for structures in isoelectronic families. Substantial progress was made by Zunger (1980) who derived dual coordinates based on pseudopotential radii achieving good separation for 565 binary compounds. This was followed by a series of studies by Villars (1983, 1984a,b, 1985) who carried out a comprehensive search for the most discerning indices and used these to rationalise thousands of binary and ternary metallic compounds. In section 3.3 we retrace these steps plotting a series of maps using current structures taken from the Inorganic Crystal Structure Database (ICSD). Finally in section 3.3.4 we discuss the work of Pettifor (1984) who eschewed purely physical meaningful quantities and instead attempted to define a scale that best separated structure types.



(a) *fcc*-Cu (b) *hcp*-Mg

Figure 3.2: Metal structures shown with octahedral (blue) and tetrahedral (green) interstices. To maximise electron density structures with only tetrahedral interstices are favoured.

Table 3.1: Some metallic radii (from Barrett and Massalski (1980))

Species	R (nm)	Species	R(nm)	Species	R (nm)	Species	R(nm)
Al	0.143	Co	0.125	Cr	0.125	Cu	0.128
Dy	0.175	Fe	0.124	Mg	0.160	Mn	0.112
Mo	0.136	Nb	0.143	Nd	0.183	Ni	0.125
Pr	0.182	Sm	0.179	Sn	0.140	Tb	0.176
U	0.138	W	0.137	Zn	0.133	Zr	0.159

3.2 Tetrahedrally close-packed structures

Electronic structure in metals Under the *free electron model* of crystalline metals, valence electrons are assumed to be completely free from their ions forming a uniformly dense, isotropic, electron gas. The cohesive energy of such a system is made up in large part by the energy of the electron gas and can therefore be maximised by atomic arrangements that maximise the electron density by efficient space filling. As a consequence many crystalline metals have high coordination numbers, high symmetry and fairly uniform electron densities.

Elemental metals commonly form face-centered cubic (*fcc*) and *hcp* structures with 12-fold coordination polyhedra while icosahedra are more common in intermetallic phases. Both *fcc* and *hcp* have one octahedral interstice or "hole" for every two tetrahedral interstices while icosahedral coordinations are exclusively tetrahedral (see fig. 3.2). Octahedral regions are not as densely packed as tetrahedral and are therefore disfavoured for maximising electron density. The existence of *fcc* and *hcp* in elemental metals can be explained by considering an icosahedron with unit radius spheres placed at the vertices such that they are all in contact with their neighbours. The remaining space at the centre can accommodate a sphere of radius 0.901. This ionic overlap costs energy which outweighs the benefits of higher electron density. Intermetallics, on the other hand, have two or more atomic sizes often with a size mismatch of less than 10% (see table 3.1 for examples) enabling the adoption of so called Tetrahedrally Close-Packed (TCP) structures containing exclusively tetrahedral interstices thereby achieving large and relatively uniform packing densities.

In the late 1950's, Frank and Kasper published a paper outlining a new way to rationalise tetrahedrally close-packed structures (Frank and Kasper, 1958). They subsequently (Frank and Kasper, 1959) used this framework to explain the existence of many known complex intermetallic structures and suggested hypothetical structures, many of which were later identified in nature. This work, along with similar analysis by D. P. Shoemaker, C. B. Shoemaker and Wilson (1957), represents one of the earliest examples of successful theoretical structure prediction by structure rationalisation.

Tetrahedral close packing At closest approach three atoms form a triangle and four a tetrahedron. As noted by Frank and Kasper a direct consequence of this is that coordination polyhedra in close-packed structures must have triangulated faces. To quantitatively analyse such polyhedra it is important to rigorously define the terms coordination number and neighbour. First, the *domain* of an atom is defined as its Voronoi cell which is the polyhedron containing all points in space that are closer to it than to any other. The concept is analogous to that of a Wigner-Seitz cell for the lattice and in fact the two are the same for monatomic crystals. A *neighbour* is thus an atom that shares a domain face with the central atom. The *coordination number* is the total number of such atoms (or equivalently the number of domain faces) and the set of all neighbour atoms is the *coordination shell*. Finally, the *coordination polyhedron* is the polyhedron whose edges join nearest neighbours in the coordination shell. An atom's domain and coordination polyhedron stand in a dual relationship to each other, however coordination polyhedra are the more natural concept to work with.

For each atom in a coordination shell we may define its *surface coordination number*, S_q, to be the number of nearest neighbours it has *in the shell* or equivalently the number of edges incident to that vertex. For example, the atoms at the vertices of a cuboctahedron are S_4, while those of an icosahedron are S_5. The set of likely surface coordinations can be reduced by simple geometric considerations. Figure 3.3 shows the set of symmetrical q-fold bipyramids where the axial length is made equal to the edge length in the ring. For equal sizes spheres faces should be as close to equilateral as possible to enable close packing. On these grounds S_3 and $S_{q\geq 7}$ can be excluded. S_5 atoms are most well suited for the task however the icosahedron has non-crystallographic symmetry and must therefore be combined with other polyhedra[1]. Between S_4 and S_6 the latter is preferred on grounds that S_4 leads to less dense octahedral interstices[2].

Euler's formula for a general polyhedron states that the V vertices, E edges and F faces are related by

$$V - E + F = 2. \tag{3.1}$$

[1] Of course Frank and Kasper were not to know about quasicrystals which are commonly found to have icosahedral tiling.

[2] C. B. Shoemaker and D. P. Shoemaker (1969) inform us that some structures with S_4 atoms can be found but with distorted octahedra while $q \geq 7$ atoms have not been found.

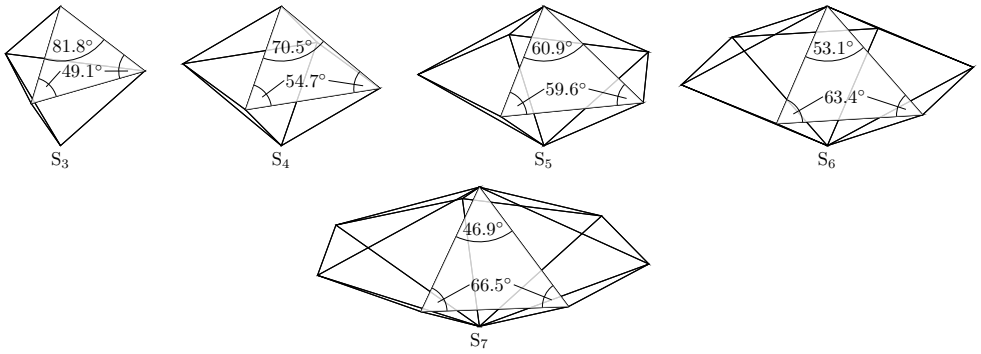


Figure 3.3: Surface coordination bipyramids in the range S_3 to S_7. Equilateral faces are most ideally suited to the formation of close-packed structures.

Defining v_q as the number of vertices with q incident edges (q-fold vertices for short), then $V = \sum_q v_q$ and $E = \frac{1}{2}\sum_q qv_q$ with the factor of one-half to account for double counting. Furthermore, for exclusively triangular faces we have $F = \frac{1}{3}\sum_q qv_q$ and thus Equation (3.1) becomes

$$\sum_q v_q - \frac{1}{2}\sum_q qv_q + \frac{1}{3}\sum_q qv_q = \sum_q (6-1)v_q = 2. \tag{3.2}$$

For polyhedra with only 5- and 6-fold vertices, and using Z for coordination number instead of V, we get

$$v_5 = 12, \tag{3.3}$$

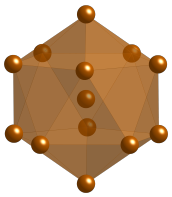
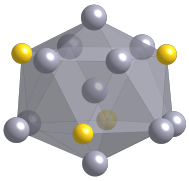
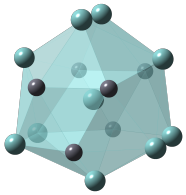
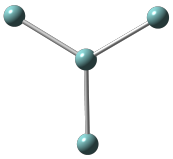
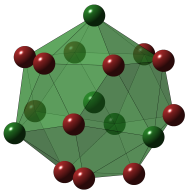
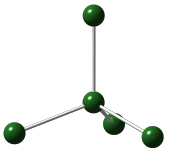
$$v_6 = Z - 12, \tag{3.4}$$

using $V = \sum_q v_q$.

An S_6 site and its neighbours are co-planar if touching equal sized spheres are used. If a neighbour is also S_6, all its neighbours will also ideally lie in the plane and thus a relatively large coordination polyhedron will be required to maintain faces that are minimally distorted from equilateral triangles. On this basis Frank and Kasper exclude coordination shells containing neighbouring S_6 sites. With this constraint there are four possible polyhedra with only S_5 and S_6 sites. The 12-coordination icosahedron and 14-, 15- and 16-coordination polyhedra detailed in table 3.2. Together these are known as the *Kasper polyhedra*. Both CN12 and CN14 have non-crystallographic point groups and are found together with one or more of the other polyhedra in TCP alloy structures.

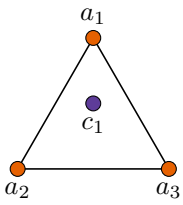
Connectivity of Kasper polyhedra In a crystal the Kasper polyhedra interpenetrate with every site at the centre of a coordination shell. To analyse their connectivity Frank and Kasper introduced the concept of a *major skeleton*. A 12-coordination site (CN12) has only S_5 atoms in its coordination shell and therefore has no neighbours with which it shares 6 common neighbours. Such sites tend to be occupied by smaller atoms and are called *minor* sites. Lines connecting two

Table 3.2: The Kasper polyhedra and their idealised geometrical properties (Sinha, 1972).

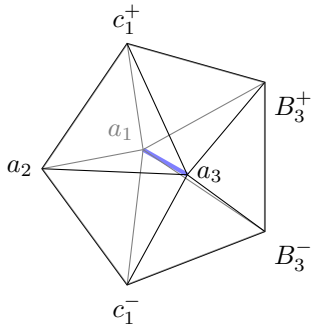
Name	Point group	v_5	v_6	E	F	Polyhedron	Major ligands
CN12	$m\overline{3}\overline{5}(I_h)$	12	0	20	20		(None)
CN14	$\overline{1}\overline{2}2m$ (D_{6d})	12	2	36	24		
CN15	$\overline{6}m2$ (D_{3h})	12	2	39	26		
CN16	$\overline{4}3m$ (T_d)	12	4	42	28		

such sites are termed *minor ligands*. Sites with $Z > 12$ (CN14,15,16) have $Z - 12$ neighbours with which they share 6 neighbours and are called *major sites*, these being occupied by larger atoms. Lines connecting major sites are termed *major ligands*. Each major site is the meeting point of $Z - 12$ major ligands which form one or more networks that extend throughout the structure, together these networks are called the major skeleton. Nodes of the major skeleton can therefore be the meeting points of 2 (in-line), 3 (in plane, 120° apart), or 4 (tetrahedrally disposed) major ligands as illustrated in table 3.2. These rules make analysis of the major skeleton substantially simpler than the structure as a whole. As pointed out by Frank and Kasper the geometry of the major skeleton is reminiscent of organic chemistry with the absence of univalent elements. This corresponds to the lack of CN13 and implies that major skeletons have infinite extent. Major skeletons with CN12-15 sites are, at least approximately, planar in the absence of large distortions. The introduction of CN16 extends the skeleton into the third dimension.

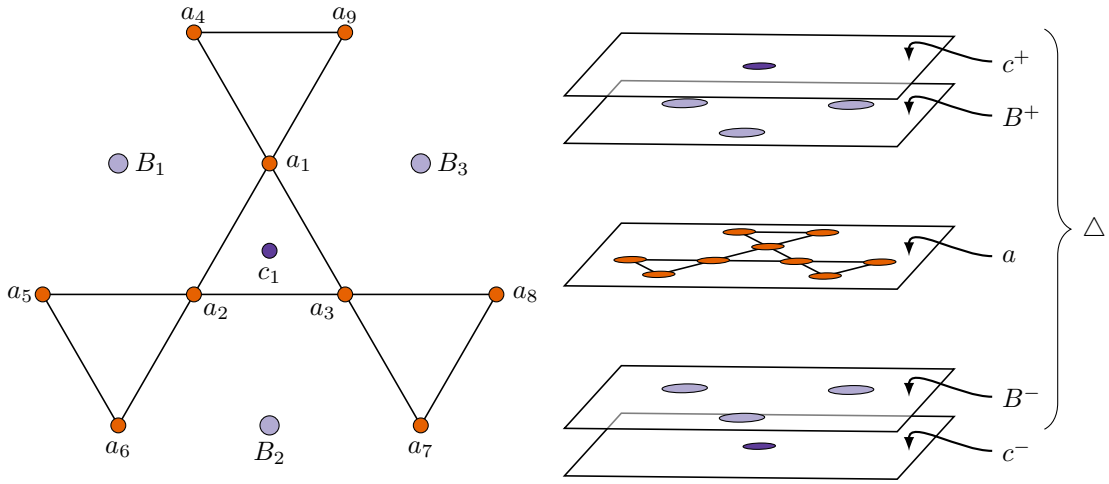
An important consequence of the rules governing the major skeleton is the occurrence of layers in TCP structures. These may be stacked in a variety of ways to build up the entire structure similar to the way AB stackings can be used to create *hcp* and *fcc* structures. By way of example let us consider the Laves phases (Laves and Witte, 1935), C14, C15 and C36. The first layer begins with a triangle $a_1a_2a_3$ which must have atoms c_1^+, c_1^- above and below to complete the tetrahedra as shown in fig. 3.4a. At most one ligand from a_1a_2, a_1a_3, a_2a_3 may be major since major ligands



(a) The starting tetrahedra. Ligands a_1a_2, a_1a_3 and a_2a_3 are minor.



(b) The corollary ring surrounding minor ligand a_1a_3.



(c) A top down and side on view of the layers.

Figure 3.4: An outline of the steps used to build the sandwich of layers, $\triangle$, that can be used to generate all three Laves structure.

Table 3.3: Examples of Frank-Kasper alloys (C. B. Shoemaker and D. P. Shoemaker, 1969)

Structure symbol	Example(s)	Pearson symbol	Space group no.	% CN12:14:15:16
A15	Nb_3Sn	cP8	223	25:75:0:0
Zr_4Al_3	$Zr_{54}Mo_{43}$	hP7	174	43:28:28:0
Fe_2B	$Fe_{67}B_{33}$	tI12	140	
C15	$MgCu_2$	cF24	227	67:0:0:33
C14	$MgZn_2$	hP12	194	67:0:0:33
C36	$MgNi_2$	hP24	194	67:0:0:33
W_2CoB_2	W_2CoB_2	oI10	71	
σ	CoCr, $Fe_{54}Cr_{46}$	tP30	136	33:54:13:0
μ	Fe_7W_6, $Co_{54}Mo_{46}$	R13	166	55:15:15:15

do not meet at acute angles. In this case we take them to all be minor. By definition a minor ligand joins two atoms that share 5 common neighbours. These form the base of a bipyramid or *corollary ring* with the minor ligand along the symmetry axis as in fig. 3.4b. To complete the ring around ligand a_1a_3 atoms B_3^+ and B_3^- are introduced and similarly for ligands a_1a_3 and a_2a_3. Finally, it remains to complete the tetrahedra associated with triangles $a_1B_1^-B_1^+$, $a_2B_1^-B_1^+$ and so on, which lie perpendicular to the plane of the page. This is done with the introduction of atoms a_4, a_5, a_6, a_7, a_8 and a_9, all coplanar with $a_1a_2a_3$. The periodic tiling of this pattern produces the Kagomé net in the a plane and triangular nets in the B and c planes. This centrosymmetric sandwich will be called $\triangle$[3], while the mirror symmetric sandwich with the central triangle pointing down will be called $\triangledown$. The C_{15} Laves phase with prototype $MgCu_2$ can be constructed by repeated $\triangle$ stackings while $\triangle\triangledown$ generates the C_{14} phase, prototype $MgZn_2$. Finally, the C_{36} phase, prototype $MgNi_2$, results from the stacking $\triangle\triangle\,\triangledown\,\triangledown$.

Frank and Kasper followed a similar procedure to that outlined above starting with various initial polygons and going through the various choices of minor and major ligands to build up sets of layers that could then be stacked to produce a variety of TCP structures. Table 3.3 shows some examples of the structures in the catalogue built up by Frank and Kasper. Along with known structures Frank and Kasper proposed new hypothetical structures some of which were later found to exist including Zr_4Al_3 and the M phase with prototype $Al_3Nb_{10}Ni_9$, a prediction that would have been challenging for even the most advanced prediction methods around today! The relevance of the work of Frank and Kasper was extended further yet with the discovery of quasicrystals in the 1980s. Shechtman et al. (1984) showed that rapid quenching of an AlMg alloy results in a structure that has point group $m\overline{3}5$ but no translational symmetry, they called this the icosahedral phase. Frank and Kasper had pointed out that of all triangulated polyhedra icosahedra are ideally suited for maximising TCP packing (fig. 3.3) but dismissed the possibility of such a structure for lack of periodicity.

[3]This definition is slightly different, but equivalent, to that used by Frank and Kasper (1959).

3.3 Structure maps

Structure maps are two or more dimensional plots of compounds showing structure type as a function of the chosen parameters. Parameters are derived from properties associated with atomic species. The goal is to identify parameters that cluster structures of common type in distinct regions achieving as few outliers as possible. Upon achieving this goal it is often reasonable to infer that a) the parameters capture the dominant structure-determining factors, and that b) the map can be used to make predictions by locating regions corresponding to compounds of unknown structure type. In principal, the greater the range of structure types that are correctly clustered the more confident we can be in the validity of these inferences. Table 3.4 shows details of a wide range of studies using different coordinate parameters achieving varying degrees of success in separating structure types and minimising outliers.

The following sections outline a series of important studies that introduced various mapping parameters aiming to correctly differentiate as broad a range of compounds and structure types as possible. We plot structure maps using a dataset retrieved from the ICSD (Allen, 2002; Belsky et al., 2002) during October 2014. In total 6615 compounds spanning 46 structure types were retrieved, this was reduced down to 5417 by excluding those with the words “metastable”, “high temperature” or “pressure” in the publication title in an attempt to retrain only low temperature and pressure, ground state structures.

Table 3.4: A list of studies using structure coordinates to map structure space (adapted from Villars (1983)).

Reference	Coordinates[a]	Compound class[b]	Structure (Number of compounds)	Total	Violations
		XY normal valence	ClNa (115), SZn_h + SZn_c (42)	157	9
Mooser and Pearson (1959)	$\overline{n}$ *vs.* $\lvert\Delta\chi^{GT}\rvert$	XZ normal valence	ClNa(33), SZn_h(5), SZn_c(7), ClCs(4)	49	3
		XT normal valence	AsNi(28), ClNa(22), MnP(22)	72	3
Pearson (1962)	$\overline{n}$ *vs.* $\lvert\Delta\chi^{P}\rvert r_X/r_Y$	XY normal valence	ClNa(115), SZn_h + SZn_c(42)	157	9
Phillips and Van Vechten (1969)	C *vs.* E_h	X^NY^{8-N}	NaCl(31), C(4), SZn_h(9), SZn_c(24)	68	-
St. John and Bloch (1974)	$\lvert(R_P+R_S)_X-(R_P+R_S)_Y\rvert$ *vs* $\lvert(R_P-R_S)_X-(R_P-R_S)_Y\rvert$	X^NY^{8-N}	C(2), ClCs(1), ClNa(27), SZn_h(8), SZn_c(24)	59	0
Machlin, Chow and Phillips (1977)		Functionally bounded suboctet X^NY^{P-N} $(3 \leq P \leq 6)$	GeK(6), NaPb(7), NaTl(7)	45	4
Watson and L. Bennett (1978)	$\lvert\Delta\chi^{WB}\rvert$ *vs.* $\overline{N}_V$	TU	AuCd(1), AuCu(14), BCr(7), ClCs(43), γ-CuTi(6), σ-FeCr(8)	85	15
Watson and L. H. Bennett (1978)	$\lvert\Delta\chi^{WS}\rvert$ *vs.* S_{XY}	Suboctet $X^{N-D}Y^D(N<8)$	AsLi(4), AsNi(2), AuCd(3), BCr(12), CdSb(2), ClCs(15), ClNa(5), GeK(6), NaPb(7), NaTl(7)	63	4
	$\Delta\chi^{WS}$ *vs.* S_X	XU	AsNi(30), BCr(18), ClCs(43), ClCs + another (15), ClNa(45), FeSi(10), MnP(22)	183	24
Zunger (1980)	$\lvert(r_p+r_s)_X-(r_p+r_s)_Y\rvert$ *vs.* $\lvert(r_p-r_s)_X-(r_p-r_s)_Y\rvert$	X^NY^{8-N} octet	C(4), ClCs(3), ClNa(65), SZn_c(29), SZn_h(11)	112	5

Table 3.4: *(continued)*

Reference	Coordinates[a]	Compound class[b]	Structure (Number of compounds)	Total	Violations
		MN non-octet	AuCd(11), AuCd(27), AsLi(4), AsNi(62), BCr(62), BFe(16), ClCs(112), ClNa(33), CoSn(3), γ-CuTi(7), GeK(6), GeS(6), FeSi(17), KO(1), LiO(3), MnP(3), MoP(4), NaHg(1), NaP(6), NaPb(7), NaSi(1), NaTl(7)	437	32
Villars (1983)	ΣVE *vs.* $\lvert\Delta(r_s + r_p)^Z_{AB}\rvert$ *vs.* $\lvert\Delta\chi^{MB}\rvert$	AB (excluding ASNi)	AsLi(6), AuCd(12), BFe(74), GeS(7), MnP(30), AlEr(12), BCr(117), AuCd(32), γ-CuTi(7), NaPb(9), CW(9), ZnS_h(21), HgS(5), AsTi(6), ClCs(298), FeSi(18), GeK(6), ZnS_c(33), ClNa(282), NaTl(14)	998	22
Villars (1984b)		AB_2	O_2V(6), FeS_2(19), $AlHo_2$(9), CV_2(6), Cl_2Pb(46), Co_2Si(13), $MoPt_2$(8), Si_2Zr(6), $CeCu_2$(55), α-$GdSi_2$(15), Cu_2Sb(20), O_2Ti(10), C_2Ca(30), $MoSi_2$(56), Al_2Cu(56), Si_2Th(20), AlB_2(102), Cd_2Ce(24), $CdIn_2$(30), $CaIn_2$(7), $CrSi_2$(10), $MgZn_2$(130), $MgNi_2$(13), FeS_2(27), CaF_2(75), Cu_2Mg(218)	1011	27

Table 3.4: *(continued)*

Reference	Coordinates[a]	Compound class[b]	Structure (Number of compounds)	Total	Violations
Villars (1984a)	ΣVE *vs.* $\lvert\Delta(r_s + r_p)^Z_{AB}\rvert$ *vs.* $\lvert\Delta\chi^{MB}\rvert$	AB_3	Se_3Zr(8), $AlCl_3$(15), β-Cu_3Ti(17), CFe_3(29), F_3Y(16), $CuTi_3$(12), $CoGa_3$(10), Al_3Ti(20), Ni_3P(8), $AsNa_3$(12), Ni_3Sn(21), Cl_3U(22), Ni_3Ti(11), D_3Ho(8), F_3La(30), Cl_3Fe(30), $AuCu_3$(253), Cr_3Si(66), As_3Co(12), BiF_3(37)	648	11
		A_3B_5	Ge_3Rh_5(11), S_5U_3(32), Nb_3P_5(5), Pd_5Pu_3(16), Pu_5Rh_3(19), B_3Cr_5(45), Si_3W_5(55), Pd_5Th_3(6), Mn_5Si_3(200)	389	1
Pettifor (1986)	$\mathcal{M}_A$ *vs.* $\mathcal{M}_B$	AB, $AB_{2-6,11-13}$, $A_2B_{3,5,17}$, $A_3B_{4,5,7}$, A_4B_5, A_6B_{23}			

[a] $\overline{n}$ is average principal valence quantum number; $\overline{N}_V$ is number of electrons per atom; $S_{XY} = S_X + S_Y$, is sum of s-p factors; S_X is s-p factor of X element; ΔX^P is Pauling electronegativity difference (Pauling, 1960); ΔX^{MB} is Martynov-Batsanov electronegativity difference (Martynov and Batsanov, 1980); ΔX^{WB} is Watson-Bennett electronegativity difference (Watson and L. Bennett, 1978; Watson and L. H. Bennett, 1978); r_X and r_Y are covalent radii of cation and anion respectively; r_s and r_p are St. John and Bloch (1974) or Zunger and Cohen (1978) pseudopotential radii; r_0, r_1 and r_2 are renormalised orbital radii; $\sum VE_{AB}$ is the sum of valence electrons of elements A and B; $\Delta(r_s + r_p)^Z_{AB}$ is the difference of Zunger pseudopotential radii sums; $\mathcal{M}$ is the Pettifor chemical scale.

[b] X and Y are non-transitional elements, Z are A group cations, T and U are transitional metals, M and N are s, p and d elements, A and B are s, p, d and f elements.

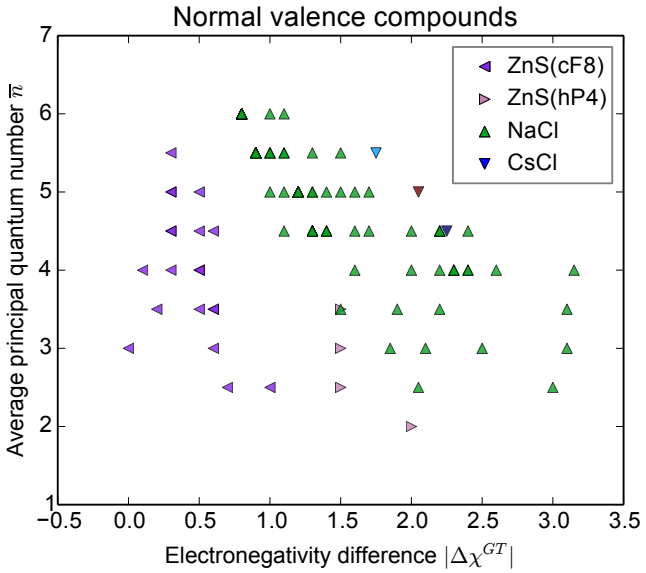


Figure 3.5: A Mooser-Pearson plot for compounds 105 'normal valence' compounds taken from the ICSD.

3.3.1 Mooser-Pearson maps

Mooser and Pearson (1959) carried out one of the first studies showing a successful use of structure maps that spanned both ionic and covalent structure types. They argue that non-directional parameters typically used to describe metallic and ionic phases (atomic radii, electronegativity differences, etc.) are inadequate when rationalising covalent structures. To account for differences in bonding character they used the principal quantum number of the valence shell, n. The reasoning behind this was expressed by Dehlinger (1955) who argued that as n increases the associated atomic orbitals become less localised and less influenced by the nucleus thereby becoming gradually less directional. For heteronuclear bonds Mooser and Pearson used the average quantum number given by $\overline{n} = \sum c_i n_i / \sum c_i$, where c_i is the number of atoms of specie i per formula unit and n_i the associated quantum number.

The second mapping parameter used is the Gordy-Thomas electronegativity (Gordy and Thomas, 1956) difference, $\Delta\chi^{GT}$. Even at fixed $\overline{n}$ bonding becomes increasingly ionic (and less covalent) as $\Delta\chi^{GT}$ increases and therefore less directional. Figure 3.5 gives an example of such a plot for AB (where A is a non-transitional metal cation) compounds forming in rock salt, caesium chloride, wurtzite (ZnS - cF8) or zincblende (ZnS - hP4) structures. The structures cluster into regions located as one may intuitively expect, going from least ionic at the bottom left, to most ionic at top right. Crucially as expressed by Mooser and Pearson the *degree of ionicity* is not governed by electronegativity alone but may be more faithfully measured by including some parameter that is sensitive to bonding directionality.

The range of compounds that are correctly clustered in Mooser-Pearson plots is somewhat limited. For example they only considered *normal valence compounds*. A generalisation of the octet

rule, the normal valence rule states that a compound, C_mA_n with e_C cation and e_A anion electrons, is normal valence if the equality $me_C = n(8 - e_A)$ is satisfied. Nevertheless this work represents an important step forward principally as it was the first to use a quantum mechanical structure parameter and therefore was able to demonstrate separation of a broader range of structure types than achieved previously.

3.3.2 Zunger maps

Zunger and Cohen (1978) build on work by St. John and Bloch (1974) by using a purely quantum mechanical description for their structure coordinates. For each element they define a first-principles pseudopotential for each angular momentum channel, s, p and d based on the DFT formalism (see chapter 2). The crossing points of the screened pseudopotentials, $V^l_{\text{eff}}(r_l) = 0$, where l is the angular momentum channel, correspond roughly to the distance away from the core at which the Pauli repulsion of the other electrons balances the inward attraction of the core and give a set of radii that mix aspects of size, electronegativity and bond type. Using the pseudopotential radii for the s and p channels, r_s and r_p, Zunger and Cohen define a set of dual coordinates for AB compounds given by

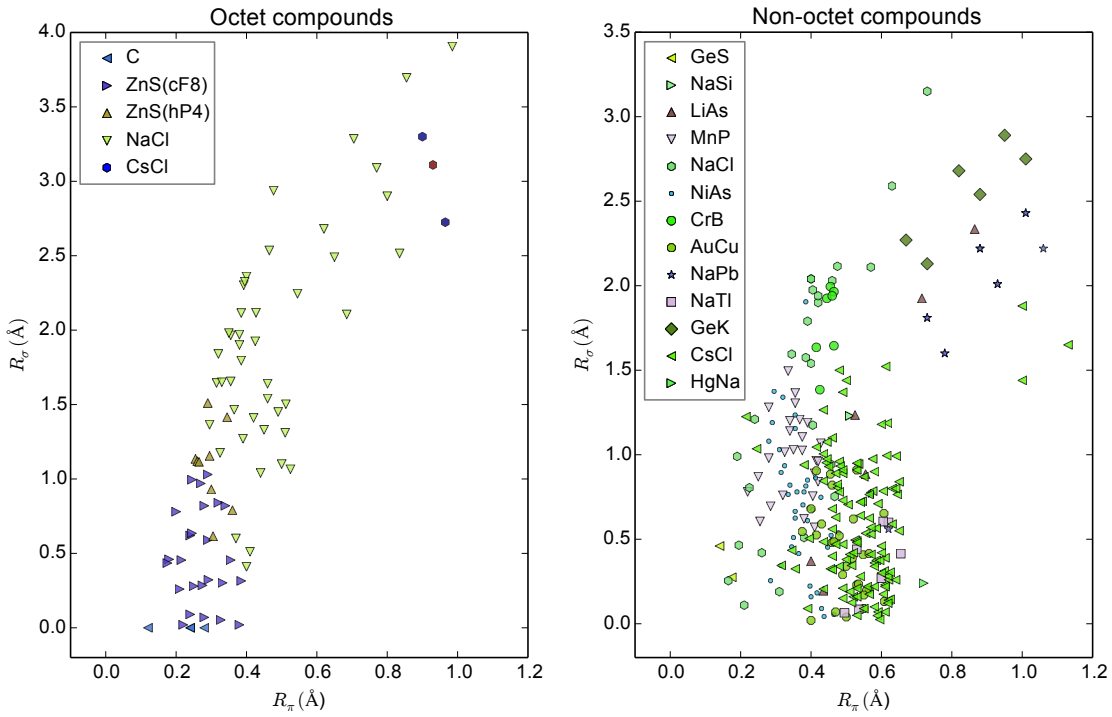
$$\begin{aligned} R_\sigma^{\text{AB}} &= |(r_p^{\text{A}} + r_s^{\text{A}}) - (r_p^{\text{B}} + r_s^{\text{B}})|, \\ R_\pi^{\text{AB}} &= |r_p^{\text{A}} - r_s^{\text{A}}| + |r_p^{\text{B}} - r_s^{\text{B}}|. \end{aligned} \tag{3.5}$$

Here, R_σ^{AB} gives a measure of the size mismatch between the two species while R_π^{AB} measures the orbital nonlocality giving a measure of bonding directionality.

Figure 3.6 shows two typical Zunger maps plotted using structures taken from the ICSD, octet and non-octet compounds have been plotted separately achieving fair separation for a range of bonding types spanning from the covalent diamond structure to the ionic NaCl structure. Just as in Mooser-Pearson plots wurtzite and zincblende structures are separated despite differing only starting from third-nearest neighbours however the range of compounds considered is now greatly expanded to encompass a large range of electronegativities, coordinations (2 to 12), ionic sizes and properties (insulators, semiconductors, semi-metals, metals and superconductors). There are structure types that cannot be separated, for example CsCl and CuAu are grouped in fig. 3.6b. Zunger points out that examples like this are often structurally very similar, as in this case which is discussed by Hume-Rothery (1950). Often compounds crystallising in two related structure types at slightly different temperatures have small formation enthalpy formation differences (see e.g. Hultgren et al. (1973)).

3.3.3 Villars maps

In an attempt to find a set of structure parameters that best separated as broad a range of structure types as possible Villars (1983) carried out a comprehensive study of the atomic parameters



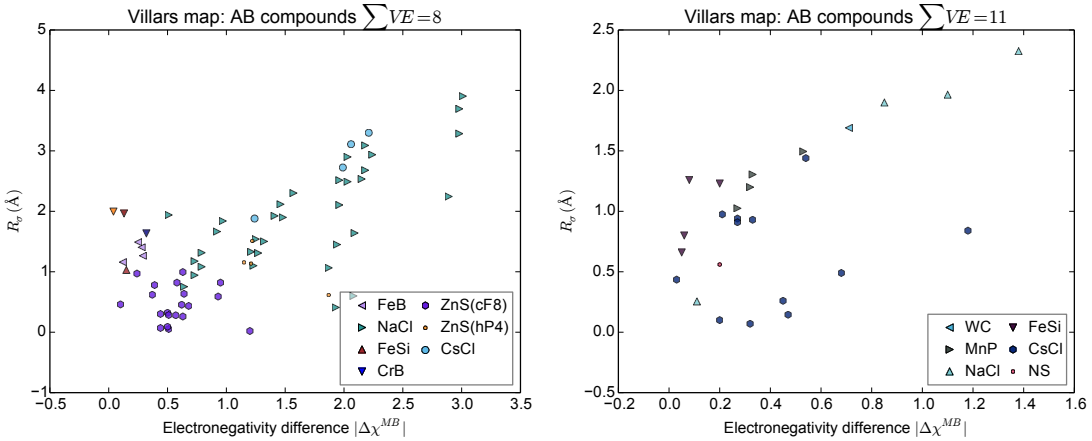
(a) 87 octet compounds.

(b) 281 non-octet compounds.

Figure 3.6: Zunger map

suggested in the literature. The goal was to find those that are minimally linearly dependent in trends across the periodic table and maximally sensitive to structure type. By plotting 182 atomic parameters against group number of each element Villars found that 5 separate patterns emerged representing parameter sets with minimal dependence. The best representatives from these were then taken and combined in various ways (sums, differences, products, etc) for compounds forming either NaCl or CsCl structures. A choice made because these structures are found across the range of ionic, metallic and covalent compounds. In the end the best set of three were found to be 1) the Martynov-Batsanov electronegativity (Martynov and Batsanov, 1980) difference , $|\Delta\chi^{\mathrm{AB}}|$, 2) the difference of Zunger's radii sums, R_σ^{AB}, and 3) the sum, $\sum VE$, of the number of valence electrons. With these in hand Villars created a series of plots featuring 998 AB compounds of 20 structure types, one plot for each value of $\sum VE$. Figure 3.7 shows two such plots for 101 structures taken from the ICSD. They both show fairly good clustering of structures into common regions with few violations.

Using the structure maps Villars produced a table of 156 predicted structure types for a range of compounds whose structure was unknown. It is interesting to look back to see those predictions that turned out to be correct. In total 17 compounds can be found in the ICSD which are correctly predicted but were published before Villars's work, with an additional 9 that were unknown at the time, these are all listed in table 3.5. Thirty seven structures are incorrectly predicted while the remainder do not form ordered structures or have no AB stoichiometry phase. Despite the number of mispredictions, the success of this result is not to be underestimated. Given the huge range of



(a) 71 AB compounds.

(b) 30 AB compounds.

Figure 3.7: Villars plots for a range of compounds taken from the ICSD at two separate value of $\sum VE$.

Table 3.5: Structure correctly predicted by Villars by locating the compounds on structure maps. In total 26 out of 156 structure were correctly predicted with many of the other not forming stable phases.

Structure type	Compound
AsLi (mP16)	AsNa
AuCd (oP4)	IrW, MoRh
BCr (oS8)	BaSi, BaGe, GeSr, LaSn, PbSr
NaPb (tI64)	NaSn
CsCl (cP2)	IrZr
NaCl (cF8)	AgC, AsNp, BiTh, CPa, CTc, CrO, CuO, KS, NPd, NTl, OPd, PSn, ZnZr
BFe(oP8) - BCr(oC8)	PdPu
GeS(oP8) - NaCl(cF8)	GeSe
AuCd (tP4) - γ-CuTi (cF8)	AgHf

factors that affect the energy landscape topology for a given compound it is quite remarkable that a set of three atomic parameters can be sufficiently sensitive to the dominant structure determining factors as to have truly predictive power. Using this new set of structure parameters Villars went on to conduct a series of studies mapping out a total of 2048 AB_2, AB_3 and A_3B_5 compounds. Overall structure clustering remained equally good with around 3% of structures falling in the wrong regions. The one major exception being the NiAs structure which Villars's excluded from his maps on account of not being able to correctly separate compounds of this type.

In the last of this series of related work, Villars (1985) attempted to derive a set of parameters that would correctly identify those systems that form stable compounds. Those not forming compounds were further categorised into one of four classes: S) solid solutions, I) insoluble, E) eutectic, and P) peritectic. Following a similar procedure to that used before, Villars compared different structure parameters to find those that were most discriminating. In the end the best set of three were: 1) the difference of Zunger's radii sums, R_σ^{AB}, 2) the magnitude of the difference in number of valence electrons, $|\Delta VE|$, and 3) the ratio, T_A/T_B ($T_A > T_B$), of the melting temperatures. The

resulting maps proved to be 96% accurate in correctly clustering compounds when applied to 3486 compounds across a range of stoichiometries. By using these maps together with the structure type maps it should be possible to predict compositional stability as well as the particular structure that will be formed by a given compound. For example it is possible to go back to the 156 predicted structures for AB compounds where one finds that 7 are correctly predicted to not from stable mixtures, thus further extending the predictive power of Villars's approach.

3.3.4 Pettifor maps

In contrast to the approaches outlined so far Pettifor (1986) rejected the idea of using parameters based on physically intuitive atomic properties instead prioritising the preservation of Mendeleev-type features across the periodic table. He argues that, for example, the failure of Villars's scheme to separate NiAs can be attributed to the fact that the parameters can give similar values for elements belonging very different groups. For instance, the Martynov-Batsanov electronegativities for IVA Zr, VIII Co and IIIB Ga are 1.70, 1.72 and 1.70 respectively. Similarly, Zunger radii show greater variation across a period than down a group. The number of valence electrons is also unable to distinguish transitional metals and p-bonded elements as it takes no account of orbital angular momentum character. Thus, Pettifor argues, use of physically intuitive parameters tends to mix up the chemical character of the elements.

To sidestep the issues with physically intuitive parameters Pettifor defines a chemical scale, $\mathcal{M}$, with no *a priori* significance, that orders the elements along a single axis while preserving Mendeleev-type features. The starting point is to sequentially order the elements going down a group, continuing up the adjacent group, back down the next, and so on. This order is then adjusted by hand to give the best possible separation of binary compounds in two-dimensional structure maps. The final order of elements can be seen in fig. 3.8.

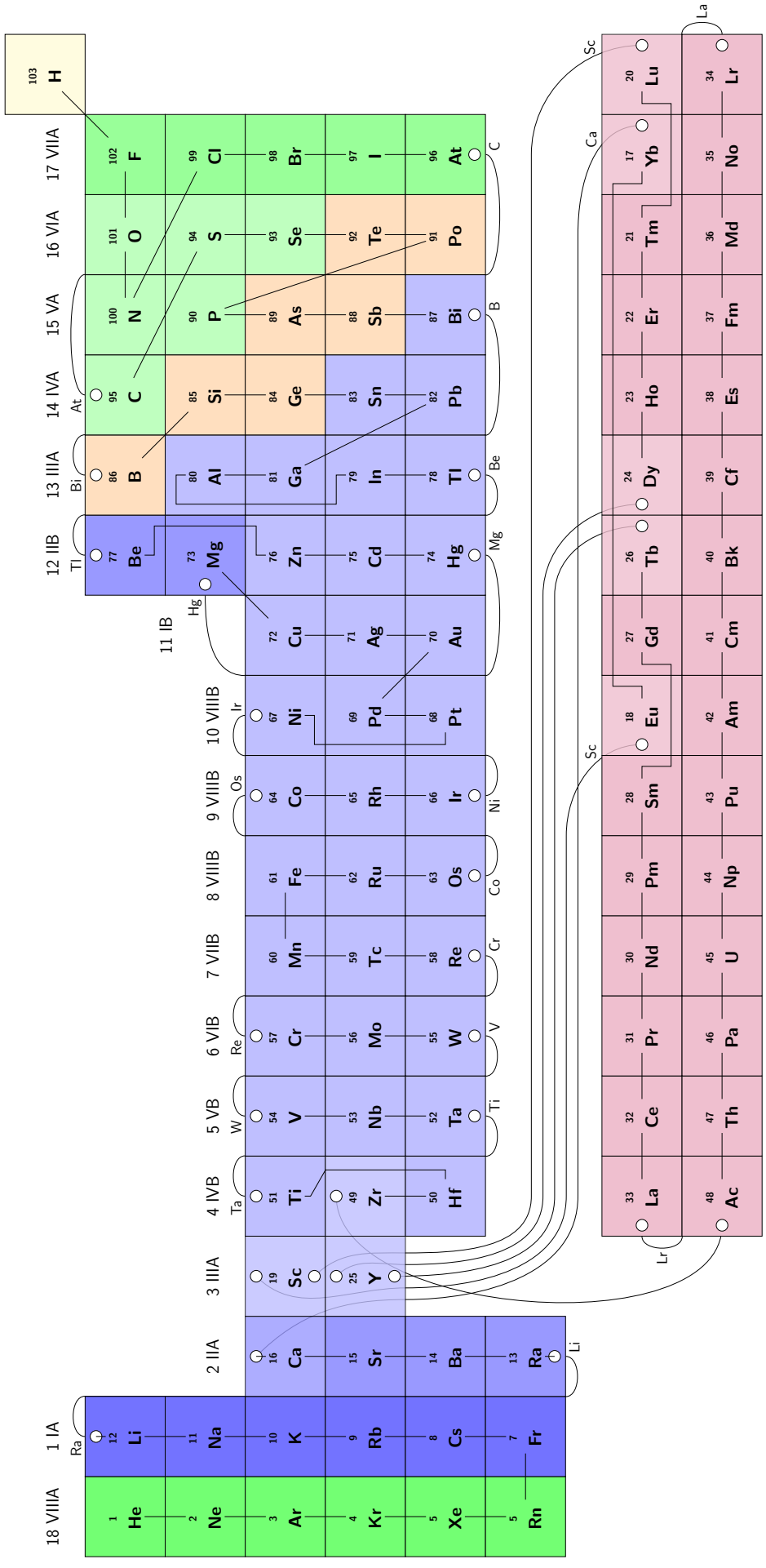


Figure 3.8: The Pettifor scale represented as a string running through the periodic table which when pulled taught gives a one dimensional scale, $\mathcal{M}$, which best separates structure types of binary compounds in two-dimensional maps.

Figure 3.9 shows a Pettifor map for the AB compounds. As expected, separation of structure types into distinct regions is very good with few outliers. The major structure types, NaCl and CsCl lie well separated in different parts of the map while even closely related structures such as FeB and CrB, NiAs, WC, TiAs and MnP, and ZnS(hP4) and ZnS(cF8) tend to cluster in distinct regions. Pettifor produced similar maps for AB_2, AB_3, AB_4, AB_5, AB_6, AB_{11}, AB_{12}, AB_{13}, A_2B_3, A_2B_5, A_2B_{17}, A_3B_4, A_3B_5, A_3B_7, A_4B_5, and A_6B_{23} achieving similar success in clustering structures. This demonstrates a robustness of the chemical scale to variations in stoichiometry.

3.4 Conclusion

The focus of this chapter has been to highlight some of the progress that has been made in understanding structure based on atomic properties that capture the structure determining portions of the cohesive energy of solids. Given that structure is ultimately determined by microscopic forces obeying the laws of quantum mechanics it appears quite remarkable that atomic properties, independent of specific environment, can be so successful in rationalising the structure types of many thousands of compounds. As pointed out by Pettifor using physically intuitive parameters may be tempting but may ultimately constrain the search for predictive parameters. Indeed more recent work in the area has moved towards computer learning on large datasets in a way that removes many human biases. Notably, Villars continues his work in the area by searching for 'materials genes' using large crystallographic databases (Kong, Luo et al., 2012; Kong, Villars et al., 2012). This enables one to build not only an understanding of individual structures but also the relationships between them.

In chapter 7 we revisit the concepts of structure rationalisation using simple parameters like size and interaction strength but rather than looking for parameters to explain sets of structures we pose the opposite question: *Given* these two parameters, what are all the structures that can occur?

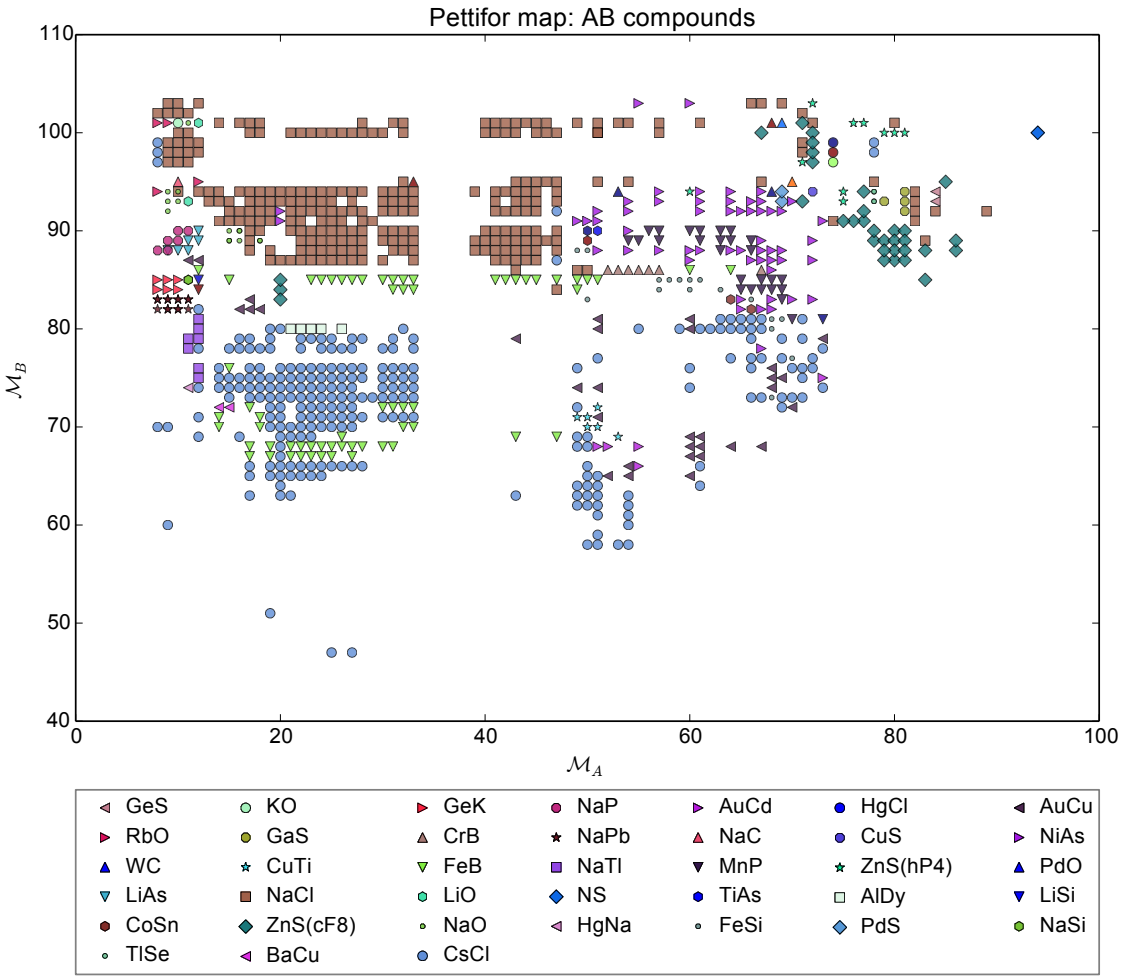


Figure 3.9: A Pettifor map of 859 AB compounds plotted showing the upper left diagonal (with $\mathcal{M}_B < \mathcal{M}_A$).

4 | Earth core intermetallics

4.1 Introduction to the inner core

The earth's interior can be divided into a series of distinct shells each characterised by a pressure and temperature range which in turn determine the characteristics of the materials that make them up (fig. 4.1). At the inner core of the Earth, where pressure range is between 330-360 GPa, sits a large spherical mass with a radius of around 860 km. Results from seismology and cosmochemistry tell us that it is composed predominantly of iron with about 10% nickel by weight (D. L. Anderson, 1989). The inner core is dynamically crystallising out of the liquid outer core, expelling lighter elements in the process. Evidence suggests that in addition to iron and nickel the inner core contains about 3% light elements (Vočadlo, 2013), the most likely candidates begin silicon, sulfur and oxygen (Alfè, Gillan and Price, 2002; Beghein and Trampert, 2003; J.-F. Lin et al., 2002) although the exact composition remains unknown. Determining the phases these adopt together with iron and nickel is an important step for interpreting seismic data which contains information about the density of the inner core.

In this work we undertake a first-principles investigation of the Fe-Ni-Si system at inner core pressures using both database structures and structure prediction to determine the phases present on the three component convex hull. Early first-principles work on the structure of FeSi at high pressure was carried out by Vočadlo, Price and Wood (1999). Under standard conditions iron monosilicide crystallises in the ϵ phase (Pauling and Soldate, 1948), however Vočadlo, Price and Wood find that at pressures greater than 13 GPa it adopts the CsCl structure. This result is mirrored by the more recent work by F. Zhang and Oganov (2010) who carried out a predictive study on the Fe-Si system using the genetic structure prediction algorithm USPEX (Glass, Oganov and Hansen, 2006). In addition to FeSi, which was found to be compositionally stable, they found a series of previously known structures, Fe_3Si, Fe_2Si, Fe_5Si_3, $FeSi_2$ and $FeSi_3$ all of which were unstable at both 300 and 400 GPa. On the Fe-Ni side high-pressure *in situ* x-ray experiments by Dubrovinsky et al. (2007) and Kuwayama et al. (2008) show that FeNi is not a stable phase while the incorporation of Ni expands the stability field of *hcp*-Fe over the competing body-centered cubic (*bcc*) and double *hcp* phases.

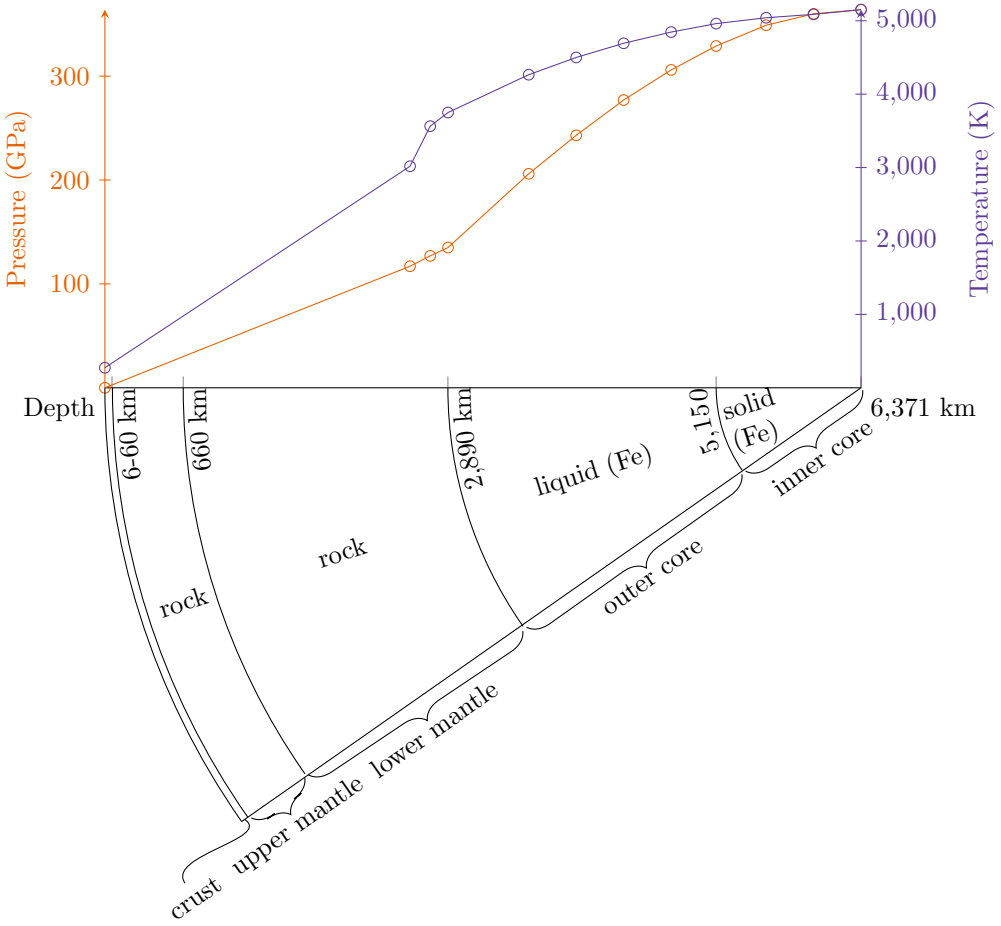


Figure 4.1: A plot of the pressure and temperature conditions as a function of depth into the earth's interior.

The Ni-Si system at high pressure has generated considerable interest. Other than the elemental structures, the NiSi composition is also stable over the relevant pressure range, adopting the MnP structure at atmospheric pressures. Under various pressure and temperature conditions NiSi can be synthesised in a fairly large range of structure types (Lord, Vočadlo et al., 2012; Lord, Wann et al., 2014) giving rise to a complex phase diagram. At 0 K recent *ab initio* calculations by Wood et al. (2013) suggest that the sequence of transitions from 21 to 264 GPa is MnP $\rightarrow$ Pmmn $\rightarrow$ CsCl although they warn that the energy differences between competing phases can be as low as 5 meV per atom. This may explain the large range of phases seen in experiment.

While both theoretical (Badro, Côté and Brodholt, 2014) and experimental (Ackerbauer et al., 2009; Antonangeli et al., 2010; Morard et al., 2011) studies have looked at the combined Fe-Ni-Si system at various temperatures and pressures up to that of the earth's liquid outer core, we know of no such studies of the solid phases up to inner core pressures.

4.2 Method

To begin populating the Fe-Ni-Si space we downloaded all the relevant structures from the ICSD and performed local relaxations at 350 GPa. The rest of the composition space was filled in using

structure prediction, concentrating on the Fe rich corner.

4.2.1 Structure prediction

High-throughput random structure searching was carried out by selecting a composition range from which random stoichiometries were automatically sampled. Each time the volume of the initial unit cell was randomly chosen around an estimate based on atomic radii and a free volume multiplier. Atoms were separated by using the following minimum separations.

Species pair	Si-Si	Si-Fe	Si-Ni	Fe-Fe	Fe-Ni	Ni-Ni
Separation (Å)	1.82	1.82	1.84	1.84	1.94	1.88

Initial structures were generated using between 2-6 symmetry operators, again sampled randomly. The volume was allowed to vary during relaxation at constant pressure.

4.2.2 Electronic structure calculations

All calculations were carried out with the CASTEP plane-wave pseudopotential code. The exact exchange-correlation energy was approximated by the PBE functional. Initial searching was performed using a plane-wave cutoff of 300 eV and a k-point spacing of 0.1. This was increased to 375 eV and 0.05 respectively for final results which was found to be sufficient to give energy differences accurate to 5 meV per atom. Spin polarisation was not treated as we expect any magnetism to be quenched at these pressures.

4.3 Results

Figure 4.2 shows a ternary plot summarising out results. The composition endpoints (i.e. elemental structure) are important as these define the chemical potentials which in turn determine the formation enthalpies for the structures on the hull. For pure iron we find the *hcp* structure to be of lowest energy, in agreement with the prevailing consensus (Côté, Vočadlo and Brodholt, 2012; Dubrovinsky et al., 2007; Kuwayama et al., 2008; Vočadlo, Price and Wood, 1999), while both nickel and silicon were found to adopt *fcc* structures with very similar lattice parameters. The rest of the phases along with their formation enthalpies are tabulated in table 4.1.

Table 4.1: Enthalpies of formation for the structures found on the hull.

Composition	Space group	Formation enthalpy (eV)
NiSi	Pmma	-1.0585
Ni_4Si_3	C2/m	-1.1426
Ni_3Si_2	Pnma	-1.1746
Ni_2Si	Pnma	-1.2004
Ni_3Si	$Pm\bar{3}m$	-1.0518
$Ni_{15}Si$	$Im\bar{3}m$	-0.2729
$FeNi_5Si_6$	C2/m	-1.2458
FeSi	$Pm\bar{3}m$	-2.1356

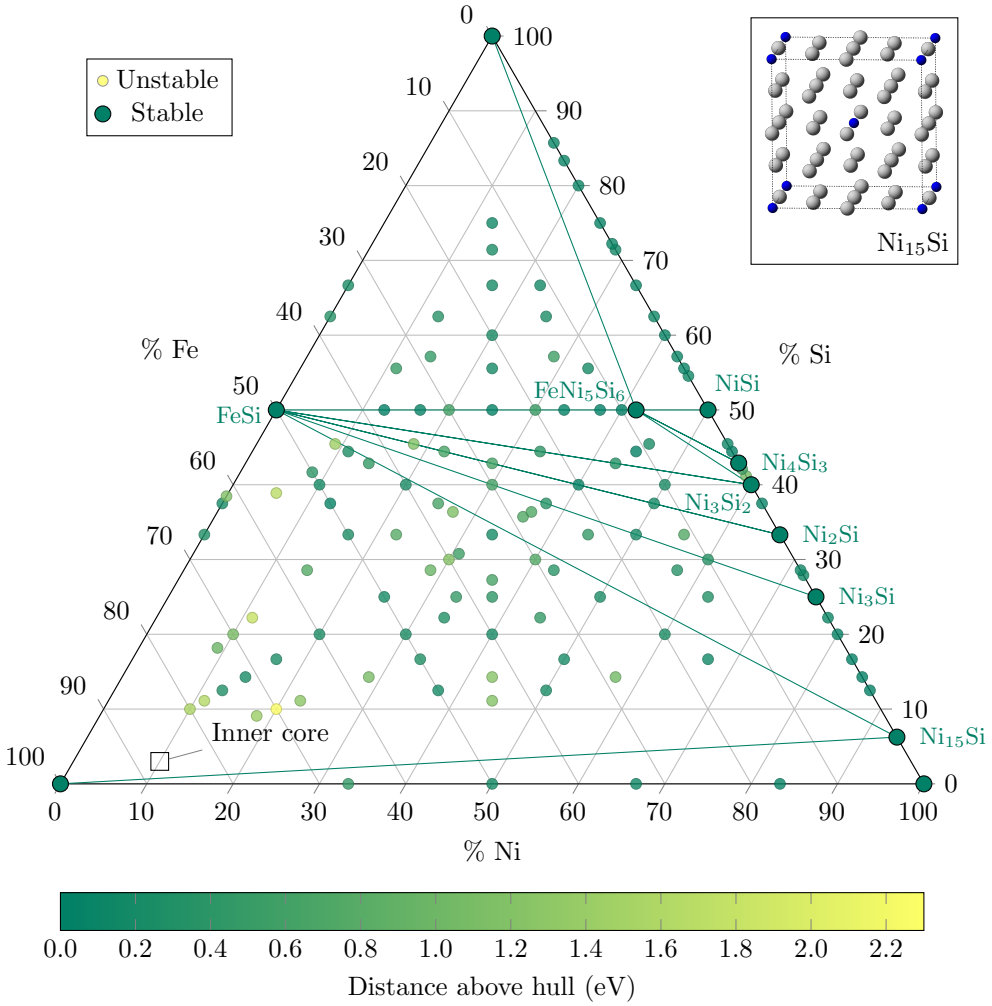


Figure 4.2: A ternary plot of the Fe-Ni-Si system at 350 GPa using both database structures and results from structure prediction.

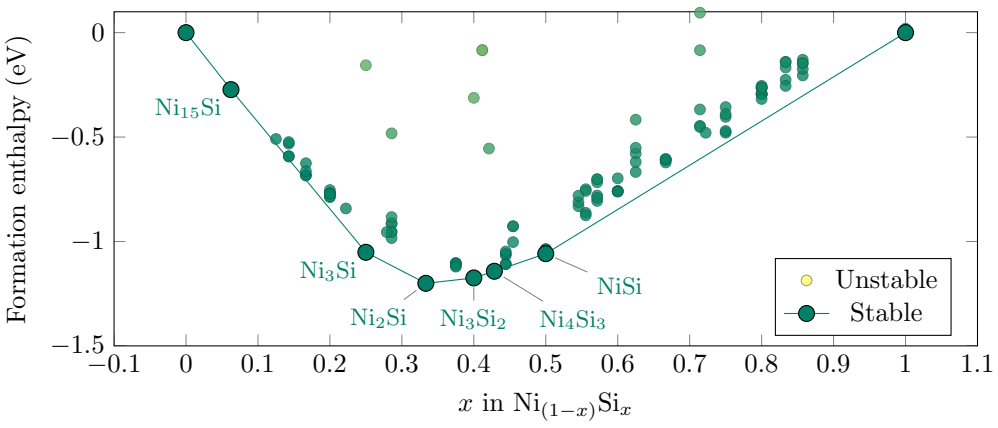
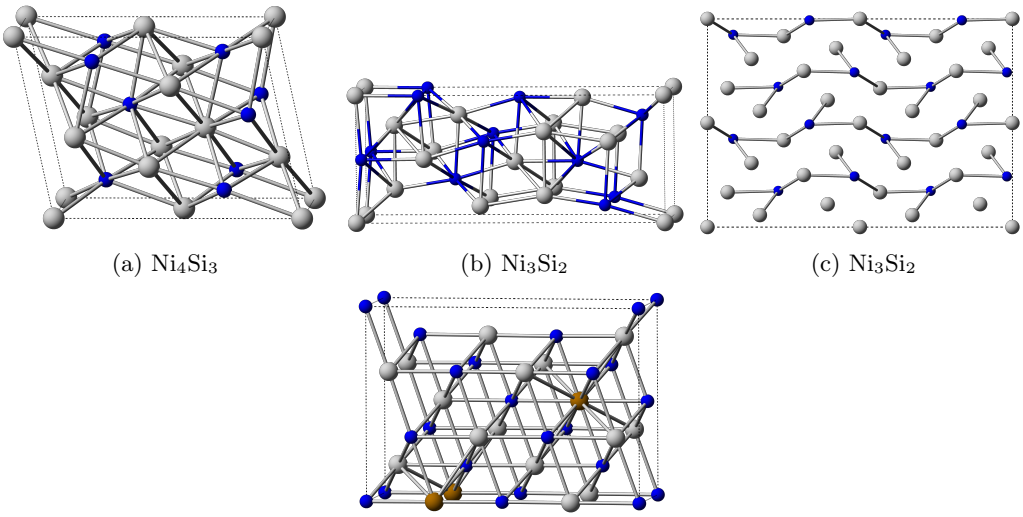


Figure 4.3: A view of the formation enthalpies for the Ni-Si portion of the system.

The most important phases for the inner core are found on the facet with vertices Fe, FeSi and $Ni_{15}Si$. As found by others previously mentioned we find that FeSi adopts the CsCl structure and that this is the only stable Fe-Si phase. More interesting is the *fcc* $Ni_{15}Si$ phase (fig. 4.2 inset) which is reported in this system for the first time. The presence of this phase on the hull displaces pure nickel as a candidate for phase segregation at inner core pressures. However, as can be seen in fig. 4.3 it protrudes only slightly along the tie line joining Ni and Ni_3Si and therefore a slight change in formation enthalpy caused by the addition of temperature could very well lift it off. Never the less, this is an important phase for future studies to consider.

4.3.1 New structures



(a) Ni_4Si_3 (b) Ni_3Si_2 (c) Ni_3Si_2

(d) $FeNi_5Si_6$

Figure 4.4: Stable Fe-Ni-Si structures

Figure 4.4 shows a series of new structures, which are found to be stable albeit in regions of the hull that are not necessarily relevant to inner core conditions. It is interesting to note the structural diversity found in the Ni-Si system as compared to Fe-Si. In the latter the other known unstable structures do not come particularly close to the hull with FeSi having a very large formation enthalpy compared to the other compositions. The difference between these two sides of the ternary hull can likely be explained by the fact that nickel and silicon both adopt *fcc* structures with lattice parameters that differ by less than 1%. This raises the possibility that nickel and silicon may form a solid solution with sites randomly populated by either of the two species. The structures in fig. 4.4 show evidence of this with layers of nickel on a square lattice populated by varying amounts of silicon.

4.4 Conclusion

We have shown the power of a combined approach of using database structures and structure prediction together to determine the structures of ternary systems, in this case Fe-Ni-Si at inner core pressures. Our results for Fe-Si and Fe-Ni are similar to those previously discussed in the literatures. The Ni-Si side of the hull shows much more structural richness largely due to the good lattice match between nickel and silicon at high pressures. The next step following on from this work is to add finite temperature to all the low lying structures as this may well change the energy ordering and indeed compositional stability. Naturally, this will not be without challenges given the extreme temperatures presents in the inner core.

Looking further ahead it is unlikely that iron, nickel and silicon alone can explain the anomalously low density of the inner core (about 10-15% less than a pure iron-nickel alloy) as determined from the propagation of seismic waves. Siebert et al. (2013) suggest that the presence of small amounts of oxygen left over from its formation could provide an explanation. This thesis is supported by the important work of Tsuno, Frost and Rubie, 2013 who performed experiments at 25 GPa and 3,000 K using a mixture of iron, iron oxide and silicate. The presence of iron oxide was found to be crucial in determining the distribution of silicon and oxygen between the silicate and the metal. Using a model derived from these results Tsuno, Frost and Rubie predict that even a modest temperature increase (from 3,000 K to 3,500 K) could bring the amounts of silicon and oxygen to above what is required to explain the observed density deficit. Very little is known on the structures of this quaternary system and this would pose an interesting challenge for a future structure prediction study.

5 | Phosphorus and arsenic

Phosphorus Remains Exciting!

Arno Pfitzner (2006)

5.1 Introduction

Phosphorus and arsenic occupy a position in the periodic table of fascinating structural diversity. As members of the group of pnictogens they have five valence electrons commonly forming covalent linkages with 3, 4 or 5 other atoms with as many as 10 neighbours possible with some kind of chemical bond. Their position within the group, between nitrogen and antimony, results in a rich bonding character enabling the formation of structures ranging from molecular crystals of pyramidal clusters to simple cubic to puckered graphene-like sheets held together by van der Waals interactions, all having similar energies. Indeed, with the possible exceptions of sulphur and boron, there are more observed phosphorus allotropes than any other element (Madelung, 2003).

Interest in phosphorus is growing to this day with theoretical predictions of multilayer and fullerene like structures (Guan, Z. Zhu and Tománek, 2014a) and a new two-dimensional material, phosphorene (Z. Zhu and Tománek, 2014). Phosphorene in particular has received a lot of attention (Guan, Z. Zhu and Tománek, 2014b; X. Han et al., 2014; Y. Liu, F. Xu et al., 2014; Xie et al., 2014) because of its interesting properties including strong elastic (Ding et al., 2014) and thermal transport (Z.-y. Ong et al., 2014) anisotropy, potential superconductivity (Shao et al., 2014) and suggestions of applications such as battery electrodes (S. Zhao, Kang and Xue, 2014).

The richness of bonding found in phosphorus and arsenic presents a unique challenge for *ab initio* structure prediction. The first comes from the sheer number of low lying structures to be explored. This is complicated by the existence of complex fibrous phases with large unit cells. The second comes from the subtle energy differences found between competing phases that require careful convergence and explicit consideration of dispersion forces. Arsenic in particular has a complex Fermi surface requiring very dense k-point grids and large energy cutoffs to capture correctly in DFT.

In this chapter we tackle some of these challenges. We begin with an *ab initio* prediction study to find the stable P_n clusters up to $n = 22$ in section 5.2.1. This is followed by a study of known and predicted bulk structures in section 5.2.2, highlighting the effects of dispersion corrections on energy ordering. Finally these structures are carefully converged in the arsenic system and compared to previous studies in section 5.3.

5.2 Phosphorus

Phosphorus was discovered in its white form by Henry Brandt sometime between 1669 and 1675 (Berger, 1996). Two further major allotropes, red and black, have been recognised for over 100 years and yet modifications of these and new structures continue to be proposed and studied to this day (J.-G. Han and Morales, 2004; Karttunen, Linnolahti and Pakkanen, 2008; Marqués et al., 2008; Pfitzner, 2006; Pfitzner et al., 2004). Electronic structure calculations have made it possible to calculate the energies and properties of these myriads of phases and yet pinning down their precise thermodynamic ordering remains difficult largely due to the small energy differences between competing phases which can be drowned out by a change in functional or the addition of dispersion corrections.

The experimental synthesis of the various forms of phosphorus often involves the application of high pressure (e.g. white or red to black) or heat (e.g. black to red) suggesting a complex energy landscape with relatively large barriers. Indeed kinetics and energy pathways between phases would appear to play a dominant role in determining the prevalence of particular allotropes and their modifications over that of simple energy ordering (Pfitzner, 2006). While this poses a challenge for experimentalists it admits the possibility that with clever synthesis routes it may be possible to create specific phosphorus structures with technologically interesting properties that remain stable for kinetic, if not, thermodynamic reasons.

5.2.1 Clusters

Introduction

The discovery of fullerenes ushered in a new era in chemistry because of their wide ranging potential for technological applications. This potential is largely a result of their enormous structural diversity (Fowler and Manolopoulos, 1995) which can be exploited to synthesise materials with special properties. Phosphorus has a similar potential for structural diversity, being able to form three bonds with a range of angles from 60° to 120° (Baudler and Glinka, 1993; Von Schnering and Hoenle, 1988). The primary impediment to similar study and possible exploitation of phosphorus has come from experimental difficulties isolating larger clusters. Phosphorus dimers and tetramers have been known experimentally for over 70 years (Jones, Ganteför et al., 1995; Jones and Hohl,

1990) and yet synthesis of larger clusters ($n \leq 24$) was only achieved little under three decades ago (T. P. Martin, 1986). It is only comparatively recently that laser ablation of red phosphorus has enabled systematic study of large clusters (e.g. $n \leq 40$ (Bulgakov, Bobrenok and Kosyakov, 2000)).

The stable structures for all clusters up to even moderate sizes ($n \leq 24$) are not known precisely owing largely to difficulties in experimental verification, yet much theoretical progress has been made. In early work Jones and Hohl (1990) used SA to search for stable clusters up to $n = 8$ and using DFT calculations were able to show that the wedge-like *cuneane* structure (fig. 5.2e) is the stable form of P_8 over the more commonly presumed O_h *cubane* form. This work was subsequently extended (Jones and Seifert, 1992) up to $n = 11$ providing several candidates at each cluster size. Following on from this Häser and Treutler (1995) carried out a comprehensive study of clusters (up to $n = 70$), polymers and bulk structures. Noting that "the identification of the most stable phosphorus clusters with 20 atoms or less already resembles a global optimization problem well beyond current computational capabilities", they instead built on work by Baudler and Glinka (1993) who outlined a set of heuristics that encode bonding rules commonly encountered between structural building blocks found in phosphorus. These rules were used to build up structure candidates the energies of which were calculated using a variety of techniques including Møller-Plesset second-order perturbation theory (Møller and Plesset, 1934), coupled-cluster theory (Purvis and R. J. Bartlett, 1982), random phase approximation and others, to give some indication of relative stabilities.

These and other studies (Böcker and Häser, 1995; Bulgakov, Bobrenok, Kosyakov et al., 2002; Bulgakov, Bobrenok, Ozerov et al., 2004; Guo, Wu and Jin, 2004; J.-G. Han and Morales, 2004; Häser, 1994; Häser, Schneider and Ahlrichs, 1992; R. Huang et al., 1995; Karttunen, Linnolahti and Pakkanen, 2008; Z.-y. Liu, R.-b. Huang and Zheng, 1996; Nava and Ahlrichs, 2008) have contributed greatly to understanding of phosphorus clusters and yet relatively little truly predictive work has been done, most studies relying on already known structures, heuristics or intuition to generate candidates. For example, J.-G. Han and Morales (2004) investigated fullerene-like clusters up to $n = 60$ finding them to be unstable with respect to dissociation into P_4 molecules. Karttunen, Linnolahti and Pakkanen (2008) attempted to define a set of structural principles for polyhedral clusters up to $n = 888$ using tetrahedral, octahedral and icosahedral cages finding tetrahedral to be the most stable across the size range. In contrast we concentrate on the smaller cluster sizes but perform a truly unbiased search for the stable clusters.

Method

In this we work we use *ab initio* random structure prediction to find the DFT ground sate clusters up to $n = 22$. All calculations were carried out using the CASTEP plane-wave pseudopotential code (Clark et al., 2005). Energy differences were converged to 2 meV per atom. This was achieved

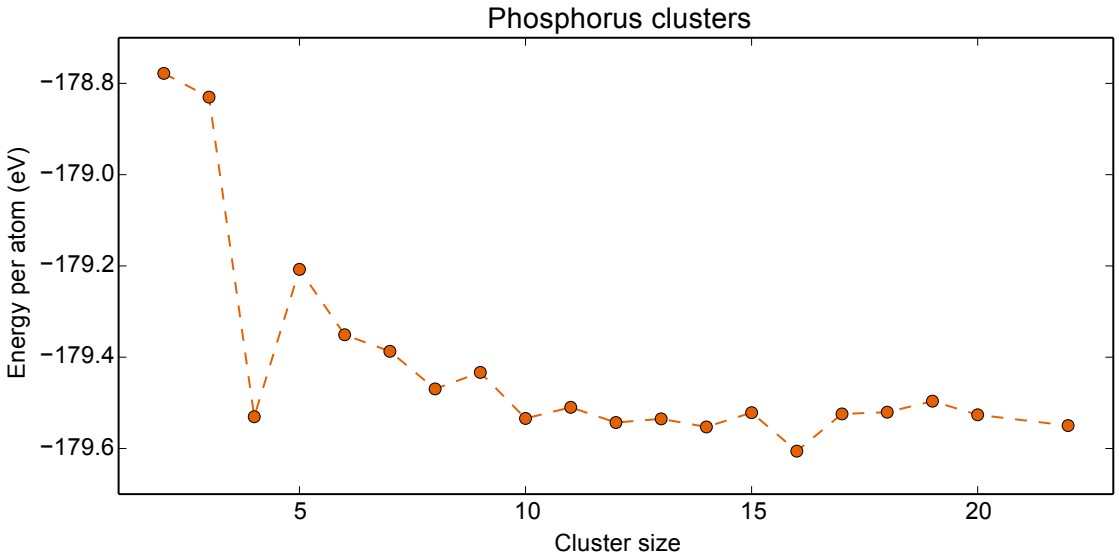
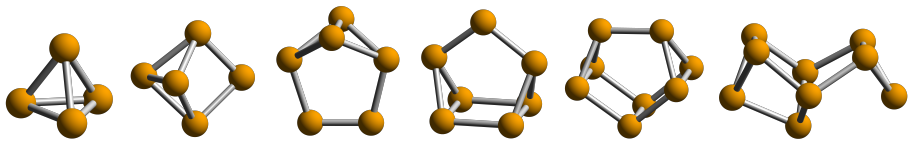


Figure 5.1: The energies of P clusters plotted for sizes up to $n = 22$. Magic sizes occur at $n = 4, 8, 10, 16$ and 18.



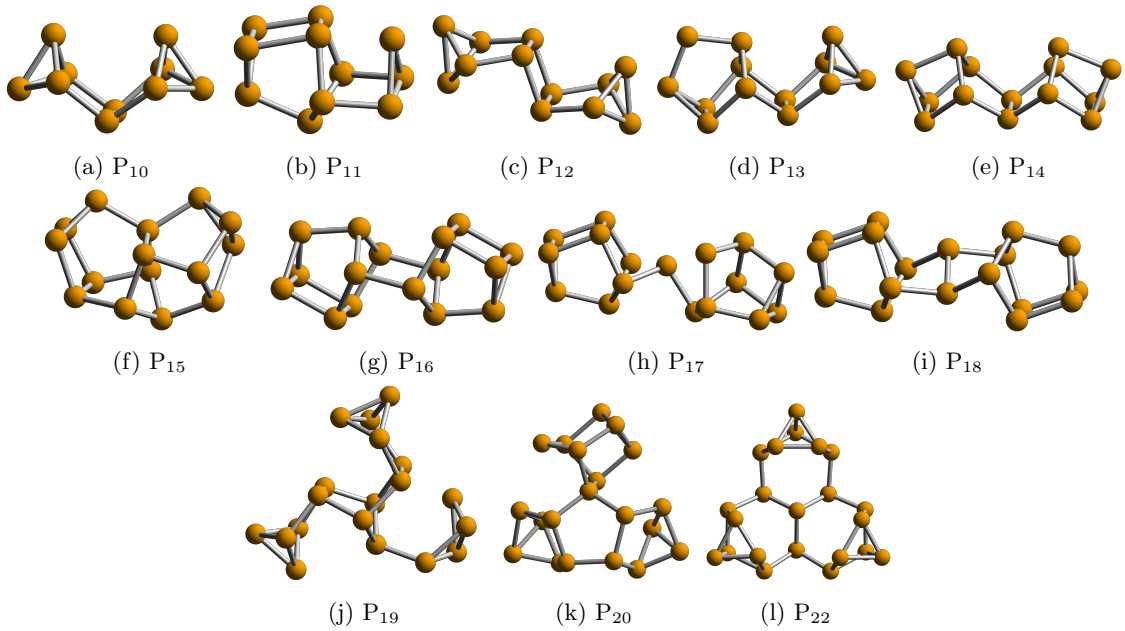
(a) P_4 (b) P_5 (c) P_6 (d) P_7 (e) P_8 (f) P_9

Figure 5.2: Lowest energy phosphorus clusters (P_n, $4 \leq n \leq 9$) found with *ab initio* structure prediction.

using a plane-wave energy cutoff of 300 eV and minimum unit cell volume of 200 $Å^3$ to mitigate any self interaction between periodic images. Calculations were carried out at the gamma point only using the PBE functional. Searching was halted when all low lying clusters had been found several times and included example of the previously published structures.

Results

Figure 5.1 shows the energies per atom for the lowest energy clusters found with images of the corresponding clusters in figs. 5.2 and 5.3. Even sizes tend to be lower in energy than the odd with magic numbers occurring at $n = 4, 8, 10, 16$ and 18. The tetramer is the unit found in white phosphorus while P_8, known as the Hittorf cluster, is the building block of violet phosphorus as discussed below. P_{10} is also found to be a "magic size" in laser ablation experiments on red phosphorus (Bulgakov, Bobrenok, Kosyakov et al., 2002). Interestingly Bulgakov, Bobrenok, Kosyakov et al. also find P_{14} to be magic which in our result does not particularly stand out, nevertheless no lower energy structures were found from over 3,000 starting configurations. A comprehensive search of the literature suggests that our clusters of sizes n = 12-15, 17, 19-22 have not been reported before.



(a) P_{10} (b) P_{11} (c) P_{12} (d) P_{13} (e) P_{14}

(f) P_{15} (g) P_{16} (h) P_{17} (i) P_{18}

(j) P_{19} (k) P_{20} (l) P_{22}

Figure 5.3: Lowest energy phosphorus clusters (P_n, $10 \leq n \leq 22$) found with *ab initio* structure prediction.

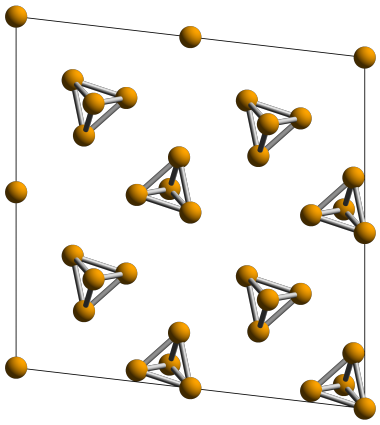
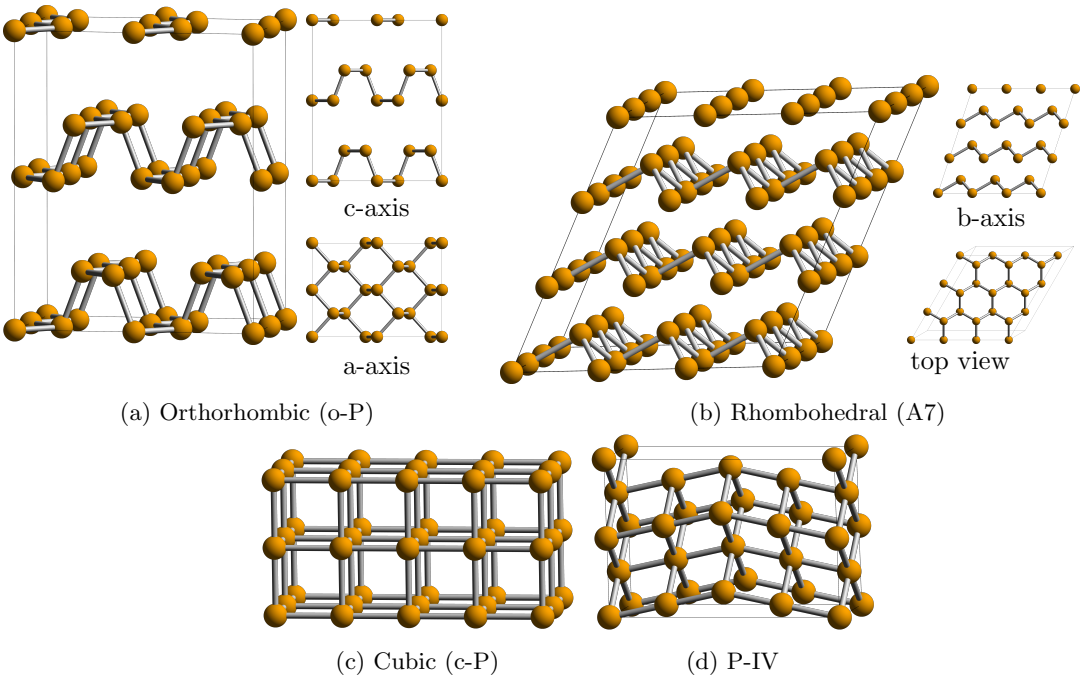


Figure 5.4: γ-white phosphorus

5.2.2 Bulk

White phosphorus

White phosphorus is a volatile, highly reactive solid that can be formed by cooling from the vapour phase. Its structure consists of tetragonal P_4 units arranged in various crystal packings. The cubic (α) form, with space group $I\bar{4}3m$, is stable at room temperature transforming into the triclinic (β) form upon cooling below 196 K. A third low temperature form (γ, show in fig. 5.4), stable at T = 123 K, is reported (Okudera, Dinnebier and Simon, 2005) to have space group C2/m with three atoms in the asymmetrical unit cell.



(a) Orthorhombic (o-P) (b) Rhombohedral (A7)

(c) Cubic (c-P) (d) P-IV

Figure 5.5: Crystalline forms of black phosphorus ordered by stability with increasing pressure.

Black phosphorus

Black phosphorus is considered to be the thermodynamically stable phase (Corbridge, 1974). Unlike white phosphorus it is largely unreactive and insoluble with highly polymeric structures. Amorphous form are reported to exist (ibid.) as well as several crystalline forms. At ambient conditions it exists in the semiconducting orthorhombic (o-P) phase with space group Cmca (fig. 5.5a). At around 5 GPa a phase transition to the semimetallic rhombohedral (A7) phase (fig. 5.5b) occurs with a further transition to the simple cubic (c-P) phase (fig. 5.5c) at 10 GPa. In both the o-P and A7 phases each atom forms three bonds to neighbours within the layer while the layers are held together by relatively weak van der Waals interactions. Compression beyond 107 GPa produces an intermediate phase known as P-IV (Akahama, Kobayashi and Kawamura, 1999) which itself transforms into the simple hexagonal structure beyond 137 GPa. P-IV has an incommensurately-modulated structure that can be best described by the superspace group Cmmm(00γ)s00 (Marqués et al., 2008). The closest three dimensional approximant has an 8 atom unit cell with space group Cmmm as shown in fig. 5.5d.

Red phosphorus

Red phosphorus describes a variety of polymeric phases that share a reddish outward appearance but have a range of densities and melting points. As with black, both crystalline and amorphous forms exists. Four crystalline forms have been reported, originally numbered II-V (I being amorphous) as their structures were not known. Since then the structures of forms IV and V have been solved. There are conflicting reports on forms II and III with some authors (Van Wazer, 1958)

suggesting that they may in fact be the same phase or even poorly crystallised varieties of form IV.

Hittorf phosphorus Thurn and Krebs (1969) solved the structure of form V, now commonly called Hittorf or violet phosphorus, in 1969. The 84 atom unit cell is monoclinic with space group P2/c. This complex phase is composed of an interlinked three-dimensional polymer network composed of double chains with pentagonal cross sections as seen in fig. 5.6. Each chain is composed of alternating P_8 and P_9 clusters joined by two additional atoms. The chains are arranged in double layers, one layer perpendicular to the other, bonded by an additional atom between P_9 clusters. Van der Waals bonding holds neighbouring double layers together.

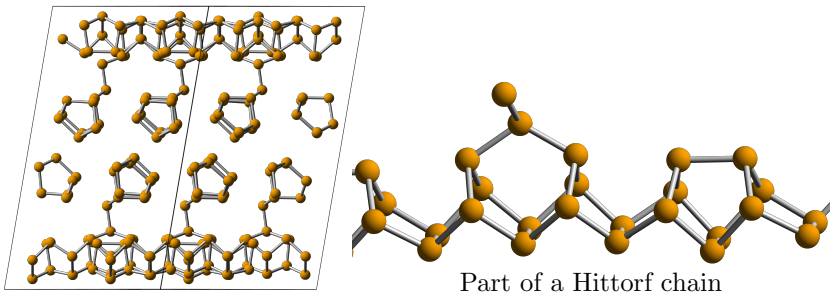


Figure 5.6: Hittorf phosphorus.

Red-IV In their experiments on Hittorf phosphorus, Thurn and Krebs also found a fibrous form of red phosphorus that was stable but less dense than Hittorf. They proposed that this form may be composed of the same chains but lying parallel to each other. Later, in their calculations, Böcker and Häser found these two structures to be isoenergetic. Finally Ruck et al. confirmed the assumed structure to be that of Red-IV using single crystal X-ray diffraction and transmission emission microscopy. The final structure, seen in fig. 5.7, has a 42 atom unit cell with space group $P\bar{1}$.

Two dimensional A7 sheet The search for a graphene-like material with a large bandgap has lead researchers to consider phosphorus as a possible candidate. Recently Z. Zhu and Tománek (2014) rediscovered the A7 phase, calling it blue phosphorus, and suggested that it may be possible to exfoliate one layer to leave a quasi two-dimensional sheet. Their DFT calculations show that this monolayer has an indirect bandgap of approximately 2 eV.

Method

Ab initio structure prediction was used to search for any new bulk phases. As for the clusters, total energy differences were converged to 2 meV per atom. As well as standard GGA calculations, Tkatchenko-Scheffler (TS) (Tkatchenko and Scheffler, 2009) and Grimme (G06) (Grimme, 2006)

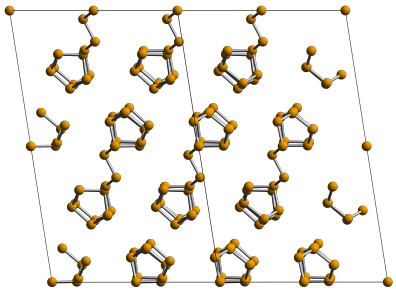


Figure 5.7: Phosphorus Red-IV. The chains are identical to those in Hittorf, except stacked parallel to one another.

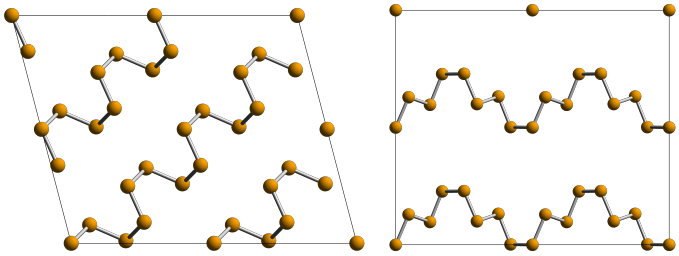
Table 5.1: Summary of the known phases of phosphorus.

Structure	Num. atoms	Space group	Also known as
Red-IV	42	$P\bar{1}$	Red phosphorus
Hittorf	84	P2/c	Violet phosphorus
o-P	8	Cmca	Black phosphorus, A17
A7	2	$R\bar{3}m$	Black phosphorus (h-P)
P-IV	2	Cmmm	
c-P	1	$Pm\bar{3}m$	Simple cubic (sc)
γ-white	8	C2/m	
β-white	24	$P\bar{1}$	

dispersion corrections were applied to assess the effect of explicit inclusion of van der Waals interactions on energy ordering. As part of the TS procedure space around ions is partitioned according to the Hirshfeld scheme (Hirshfeld, 1977) which can cause spurious results in small unit cells if regions overlap with themselves. In these cases energies were further converged with respect to increasing supercell size. All calculations were carried out at zero pressure.

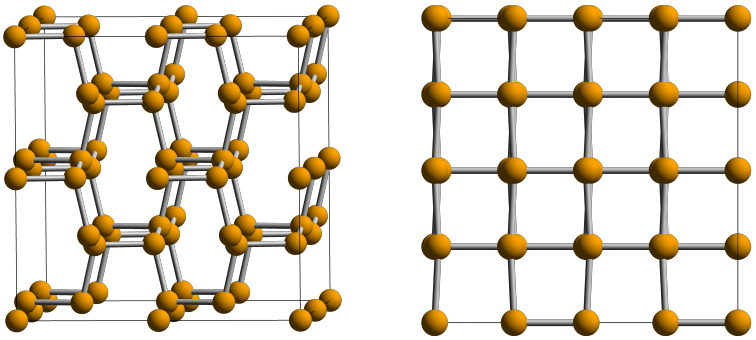
Results

Structure prediction identified a number of low energy modifications of the black phosphorus phases. Figure 5.8 shows two of the o-P modifications found, the first (o-P-mI) exhibits a stacking defect while the second (o-P-mIII) is defective within each layer. Figure 5.9 shows two of the modifications to the c-P structure both showing degrees of variation from perfect cubic nearest-neighbour bond lengths. c-P-mI in particular is severely distorted resembling something between the c-P and o-P phases.



(a) o-P modification I. (b) o-P modification III.

Figure 5.8: Two modifications of the o-P black phosphorus structure found during structure prediction.



(a) c-P modification I. Nearest-neighbour bond lengths are severely distorted from the ideal 2.43 Å found in the cubic structure, ranging from 2.23 to 3.20 Å

(b) c-P modification II. Nearest-neighbour bond lengths vary between 2.3 to 2.53 Å.

Figure 5.9: Two modifications of the c-P black phosphorus structure found during structure prediction.

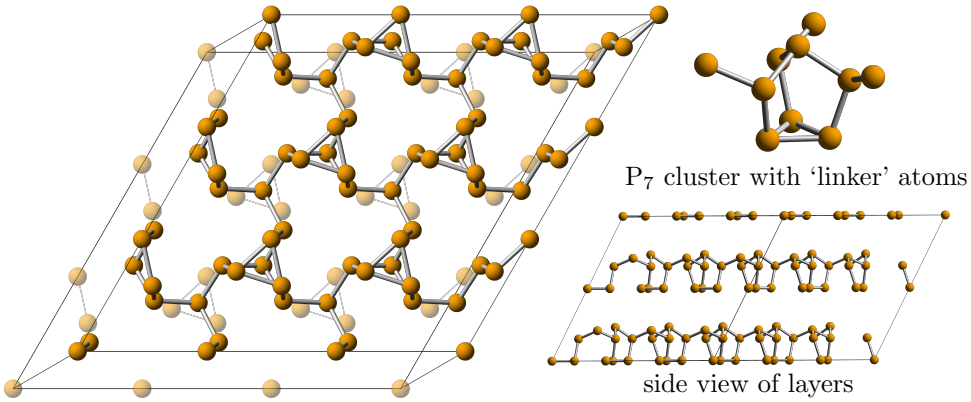


Figure 5.10: Previously unreported P_7-cl structure.

One previously unreported structure (P_7-cl), shown in fig. 5.10, was found consisting of P_7 clusters on a two-dimensional triangular lattice, each bonded to its neighbours by a single additional atom. These sheets are then stacked with layers held together by van der Waals interactions.

The dispersion corrected energies relative to GGA for the various phases considered are plotted in fig. 5.11 with further information including absolute energy orderings tabulated in table 5.2. Orthorhombic black phosphorus is not found to be lowest in energy in any of the sets of calculations, the addition of dispersion corrections does little to bring the energy ordering in line with experiment. A possible reason for this is discussed below. In all cases, however, o-P is found to be lower in energy than any of its modifications. In contrast this is only true for the A7 phase with dispersion correction while the c-P phases are generally fairly spread in energy, particularly when using G06 corrections. The newly found P_7 cluster phase falls between the red and the white phosphorus phases in energy.

Our GGA results are in agreement with similar work by Bachhuber et al. (2014) showing the same order and approximate energy differences between the various structures. This is also true

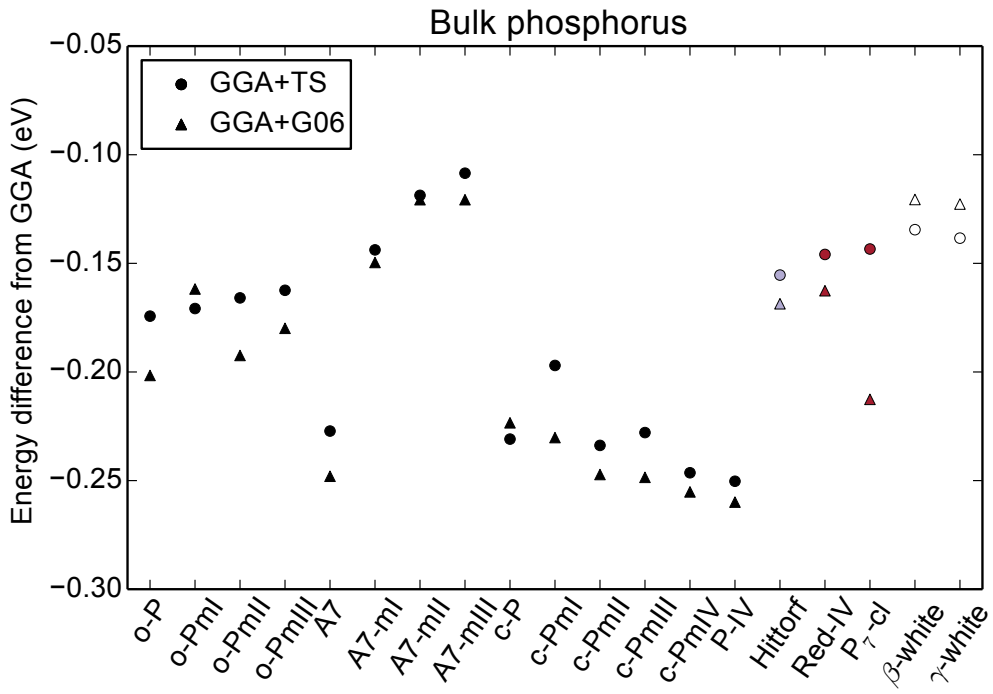


Figure 5.11: A plot of dispersion corrected energies relative to GGA for bulk phases of phosphorus for both known structures and those found using structure prediction.

for entries found in the Materials Project database[1] (Jain et al., 2013) carried out using the VASP plane-wave pseudopotential code. Bachhuber et al. also applied G06 dispersion corrections. G06 and TS are very similar differing only in their dispersion coefficients and damping function, which in TS are charge-density dependent allowing it to adapt the van der Waals contributions to local chemical environment. Using G06 Bachhuber et al. calculate the energetic ordering to be: o-P, Hittorf, Red-IV, A7, c-P. This is quite different from ours even when using the same corrections, and closer to what one would expect based on experimental evidence. It is unclear at this stage what the source of the discrepancy is.

5.3 Arsenic

Because of their adjacency in the periodic table arsenic crystallises in similar structures to phosphorus. In this section we use the structures from the phosphorus system to calculate energies and relative orderings in arsenic. The most comprehensive work of this type was carried out by Osters et al. (2012) who similarly calculated DFT energies using both nitrogen and arsenic as substituents in the known phosphorus structures. They also attempted to synthesise the o-P structure in arsenic but were unable to do so without the presence of some phosphorus. In their calculations this structure is marginally metastable over the known arsenic ground state, however they failed to taken into account any consideration of dispersion forces.

[1] Accessed on the 26$^{\text{th}}$ of November, 2014

Table 5.2: Phosphorus structures ordered by relative energy per atom using, GGA, GGA+TS and GGA+G06 dispersion corrections.

GGA				GGA+TS				GGA+G06			
Structure	Vol./ atom	Energy/ atom (eV)	Rel. energy/atom	Structure	Vol./ atom	Energy/ atom (eV)	Rel. energy/atom	Structure	Vol./ atom	Energy/ atom (eV)	Rel. energy/atom
Red-IV	27.14	-179.6914	0	A7	14.84	-179.8552	0	A7	14.97	-179.8760	0
Hittorf	23.83	-179.6874	0.0040	c-PmII	14.04	-179.8552	0	c-PmII	14.25	-179.8686	0.0073
o-P	20.93	-179.6604	0.0310	c-PmIV	14.02	-179.8527	0.0025	c-PmIII	15.00	-179.8668	0.0091
o-PmI	21.57	-179.6560	0.0354	c-P	13.81	-179.8503	0.0049	o-P	18.94	-179.8620	0.0139
o-PmII	22.03	-179.6535	0.0379	P-IV	13.81	-179.8471	0.0082	c-PmIV	14.11	-179.8615	0.0145
o-PmIII	30.33	-179.6460	0.0454	c-PmIII	14.73	-179.8462	0.0090	P-IV	14.01	-179.8567	0.0193
A7-mI	32.92	-179.6428	0.0486	Hittorf	21.96	-179.8428	0.0125	Hittorf	21.55	-179.8558	0.0193
A7-mII	27.94	-179.6418	0.0496	Red-IV	22.20	-179.8373	0.0179	Red-IV	21.25	-179.8540	0.0219
A7-mIII	33.58	-179.6407	0.0507	o-P	19.54	-179.8347	0.0206	c-PmI	14.02	-179.8526	0.0234
P_{7}-cl	26.92	-179.6374	0.0540	o-PmI	19.68	-197.8268	0.0284	P_{7}-cl	19.38	-179.8500	0.0260
A7	15.55	-179.6280	0.0634	o-PmII	20.16	-179.8194	0.0358	o-PmII	19.75	-179.8459	0.0301
c-PmI	19.06	-179.6224	0.0690	c-PmI	19.68	-179.8194	0.0358	c-P	14.03	-179.8428	0.0332
c-PmII	14.24	-179.6214	0.0700	o-PmIII	20.39	-179.8084	0.0468	o-PmIII	19.44	-179.8259	0.0501
c-P	14.19	-179.6194	0.0720	A7-mI	21.58	-179.7866	0.0686	c-PmI	18.50	-179.8178	0.0581
c-PmIII	15.36	-179.6183	0.0731	P_{7}-cl	23.45	-179.7808	0.0744	A7-mI	22.85	-179.7924	0.0835
c-PmIV	25.83	-179.6063	0.0851	A7-mII	23.85	-179.7605	0.0947	A7-mII	24.07	-179.7625	0.1135
P-IV	14.36	-179.5968	0.0946	A7-mIII	29.30	-179.7492	0.1060	A7-mIII	28.29	-179.7614	0.1146
γ-white	34.55	-179.5527	0.1387	γ-white	27.36	-179.6911	0.1641	γ-white	24.59	-179.6754	0.2005
β-white	34.94	-179.5525	0.1389	β-white	27.75	-179.6870	0.1682	β-white	24.91	-179.6731	0.2028

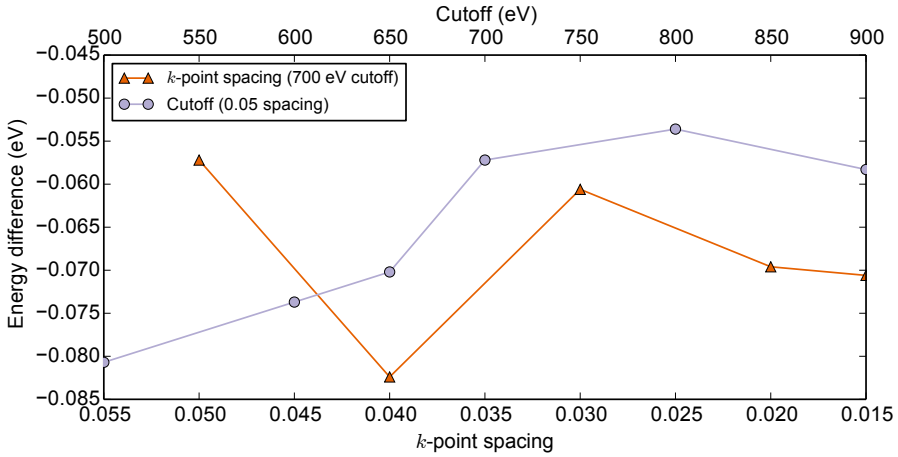


Figure 5.12: Convergence of the energy difference between o-P and c-P arsenic structures. A plane wave cutoff of 700 eV and k-point spacing of 0.02 were deemed to be sufficient to give energy differences accurate to 2 meV per atom.

5.3.1 Structures

At normal pressure arsenic is known to crystallise in three phases. The stable phase, so called gray arsenic (or α-As), exists as a semimetal in the A7 structure (Norman, 1998). Yellow arsenic has cubic symmetry and is presumed to consist of As_4 tetrahedral in analogy with white phosphorus phases (Madelung, 2003). It decomposes into grey arsenic easily and has not been structurally well characterised as it is destroyed by X-rays. There are conflicting reports on the structure of the final modification, black arsenic (or ϵ-arsenic). It is variously reported to exist in the o-P (black phosphorus) structure (ibid.) or in the red phosphorus structure (N. Wilberg, Holleman and E. Wilberg, 2001). Under the application of pressure gray arsenic is known to transform to the simple cubic structure at around 28 GPa (Silas, Haynes and Yates, 2013). To avoid confusion we shall use the structure labels used for phosphorus above (table 5.1) throughout.

5.3.2 Fermi surface

Ever since the work of P. J. Lin and Falicov (1966) arsenic has been known to have a complex Fermi surface, particularly in its ground state α-As form. The more recent efforts of Silas, Haynes and Yates (2013) confirm that to converge this system properly it is necessary to use high plane-wave cutoff energies and dense k-point grids. This is particularly important when attempting to capture small energy differences to determine correct energy ordering of competing phases. Figure 5.12 shows our convergence testing to achieve energy differences accurate to 2 meV per atom for which a plane-wave cutoff of 700 meV and k-point spacing of 0.02 were found to be sufficient.

5.3.3 Results

Figure 5.13 shows the energies of the structures considered with further data in table 5.3. The sequence of structure without dispersion corrections is different from that found by Osters et al. They get: A7, o-P, Hittorf. It is difficult to pinpoint the source of this difference as they do not provide adequate details of their calculation method. In this case the TS corrections improve the energy ordering bringing the experimental ground state, A7, to be lowest in energy followed by two of the red phosphorus phases.

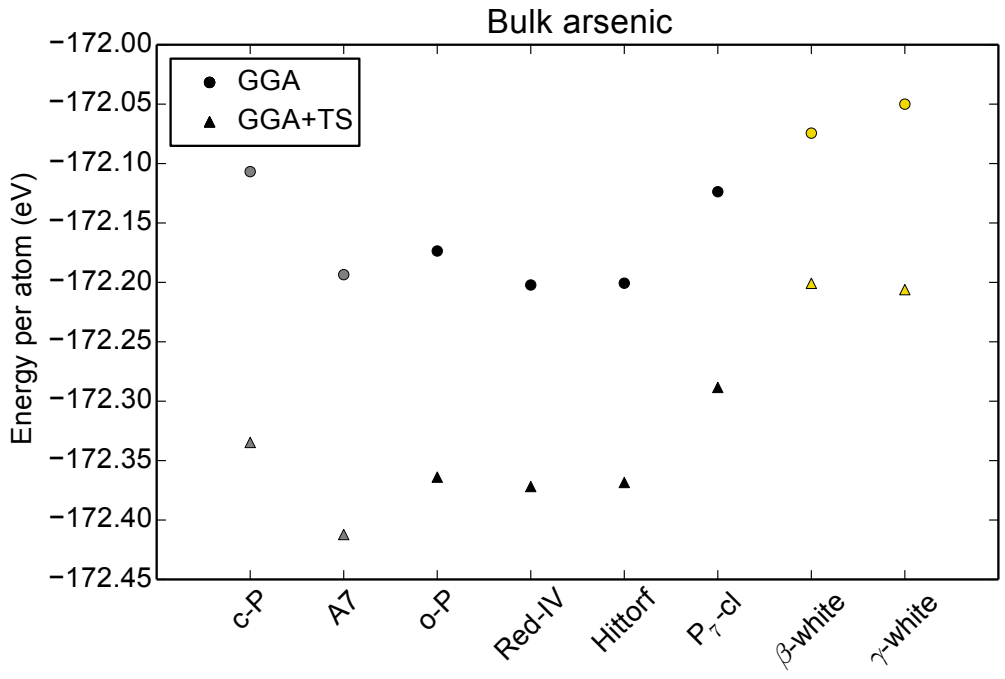


Figure 5.13: A plot of relative energies for bulk phases of arsenic. Two series are plotted, one with (triangles) and one without (circles) dispersion corrections.

Table 5.3: Arsenic structures ordered by relative energy per atom using both GGA and GGA+TS dispersion corrections.

GGA				GGA+TS			
Structure	Vol./ atom	Energy/ atom (eV)	Rel. energy/atom	Structure	Vol./ atom	Energy/ atom (eV)	Rel. energy/atom
Red-IV	28.92	-172.2022	0	A7	21.41	-172.4122	0
Hittorf	28.98	-172.2007	0.0015	Red-IV	26.52	-172.3718	0.0405
A7	21.92	-172.1935	0.0087	Hittorf	27.67	-172.3684	0.0438
o-P	24.89	-172.1736	0.0286	o-P	23.51	-172.3639	0.0484
P_7-cl	34.24	-172.1237	0.0785	P_7-c	27.97	-172.2884	0.1238
c-P	13.81	-172.1068	0.0954	c-P	13.76	-172.3347	0.0775
β-white	33.83	-172.0744	0.1278	γ-white	29.82	-172.2061	0.2062
γ-white	32.46	-171.9658	0.2364	β-white	31.10	-172.1690	0.2432

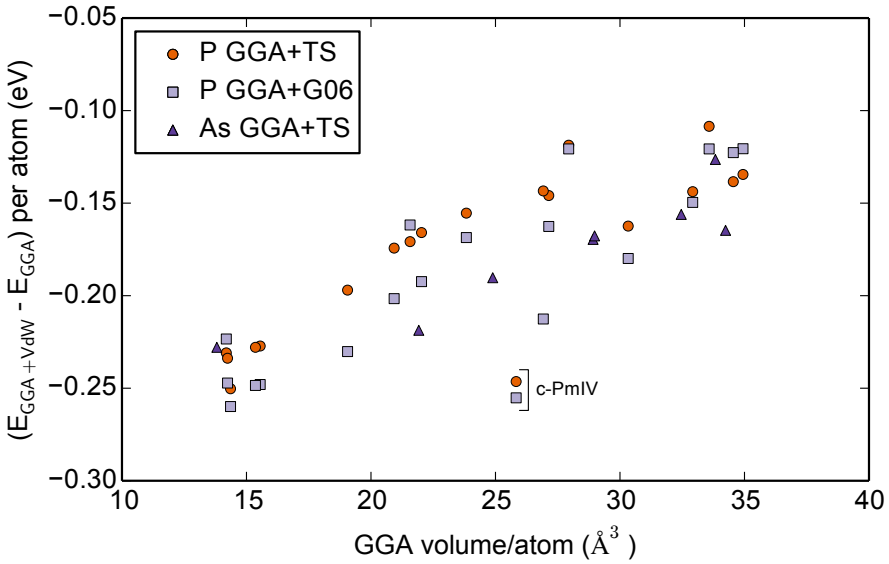


Figure 5.14: A plot showing the energy difference in switching on Grimme (phosphorus only) and Tkatchenko-Scheffler dispersion corrections as a function of volume per atom in bulk phosphorus structures and arsenic.

5.4 Discussion of dispersion corrections

Dispersion corrections have a rather significant effect on the energy ordering of phosphorus phases in particular. This is perhaps not surprising giving the range of bonding encountered. The o-P phase is particularly affected by the addition of TS corrections, rising to 20 meV per atom higher in energy than the A7 high-pressure phase. A plot of the energy difference in switching on dispersion corrections against volume per atom is shown in fig. 5.14. A trend whereby denser phases show a greater decrease in energy can be seen showing the source of much of the differences between the sets of calculations. The trends are rather close for both G06 and TS although perhaps more tightly clustered along a line in the case of TS. Both correction schemes show c-PmIV to be an outlier in being particularly affected by the corrections. This may prove to be a useful test case for comparing various dispersion correction schemes.

Given that there is significant disagreement between all three sets of phosphorus calculations these results suggest that the current set of semi-empirical dispersion corrections cannot be considered to give accurate energy orderings for complex bulk systems that demonstrate a wide range of bonding characters and small energy differences between phases. The next step in a more detailed investigation would be to try more sophisticated correction schemes that modify the electron density directly through the functional such as DFT-DF (Dion et al., 2004) or optB88-vdW (Klimeš and Michaelides, 2012).

5.5 Conclusion

The phosphorus and arsenic systems continue to provide an interesting challenge for structure prediction. We have found one new phosphorus structure that is within the energy range of the known phases and it is not inconceivable that others exist that are yet to be identified. The kinetic accessibility of these phases is, however, an open question and could be an impediment to synthesis. Furthermore questions have been raised about the accuracy of DFT calculations in situations such as this where van der Waals forces can seemingly play the deciding role in energy ordering because of the small energy differences between competing phases. Nevertheless our work has brought together a large amount of previous knowledge on the two systems in a consistent way and should act as a good starting point for those investigating these and other outstanding questions.

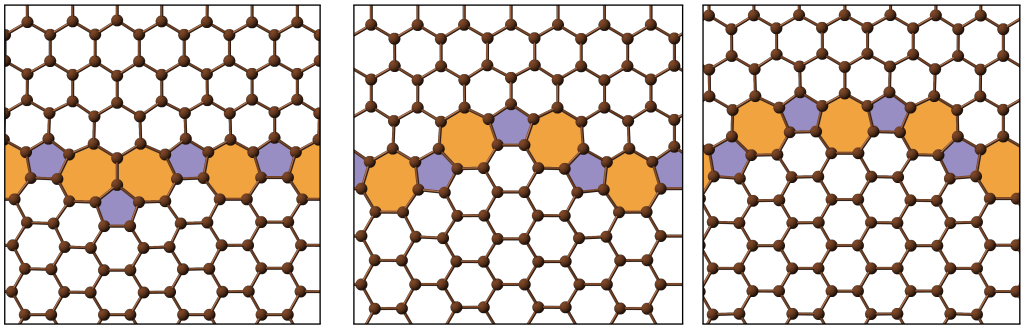
6 | 2D materials interface prediction

6.1 Introduction

The successful synthesis of graphene by Novoselov et al. (2004) sparked a 2D materials revolution. Research output on graphene alone has reached almost ten thousand papers annually and attention has begun to focus on a range of other 2D crystals (Butler et al., 2013; Mas-Ballesté et al., 2011; Osada and Sasaki, 2012; Wang et al., 2012; Yong Xu et al., 2013). Collectively, the draw of these materials is often the promise of exiting and unusual mechanical, electrical or optical properties. In practice these properties are sensitive to fabrication defects which must be taken into account if this potential is to be delivered upon in real applications. The most common technique, Chemical Vapour Deposition (CVD), unavoidably results in polycrystalline samples (An et al., 2011; Duong et al., 2012; P. Y. Huang et al., 2011; Kim et al., 2011; Yu et al., 2011) as the result of multiple nucleation sites which grow and coalesce into domains divided by grain boundarys (GBs). The GBs are typically long lived and cannot be annealed out as the energy cost of rotating an entire grain is large. Their presence gives rise to a long-range stress field (Warner et al., 2012; Wei et al., 2012) and affects the electrical (Cummings et al., 2014; Fei et al., 2013; Koepke et al., 2013; Kou et al., 2011; T. I. N. G. Li et al., 2012; S. S. Roy and Arnold, 2013; Tapasztó et al., 2012; Tsen et al., 2012; Yazyev and Louie, 2010), optical (Fei et al., 2013) and mechanical (Grantab, Shenoy and Ruoff, 2010; Song et al., 2013; Wei et al., 2012; J. Zhang, J. Zhao and Lu, 2012) properties of the crystal. To be able to accurately model these properties it is therefore necessary to know the atomic arrangements adopted at the grain boundaries.

Inevitably nature will adopt the lowest energy arrangement for a given tilt angle, θ, between two grains. Several studies using various optimisation approaches that have attempted to identify the stable configurations for a range of tilt angles. Y. Liu and Yakobson (2010) were one of the first to tackle the problem building interfaces by inserting pentagons and heptagons at various concentrations along the boundary. Z.-L. Li et al. (2014) build on this by using a genetic algorithm to perform global optimisation on the armchair/zigzag interface. Both these studies employed

empirical potentials when searching for configurations, in the case of Z.-L. Li et al. the best few were relaxed in DFT. However, empirical potentials are prone to larger energy errors at grain boundaries where conditions are often non-stoichiometric and neighbour environments and coordinations are away from what the potential was fitted to. Recent work of Schusteritsch and Pickard (2014) highlights this shortcoming. They performed random structure prediction on the $\theta = 30°$ armchair-zigzag boundary and found an additional low-lying configuration, GB-III, missed in previous work. Figure 6.1 shows the all three GBs. Depending on if the constrained cell length along the interface is constrained to ideal armchair or zigzag, GB-II or -III become the lowest energy with GB-I lying somewhat higher in energy.



(a) GB-I (b) GB-II (c) GB-III

Figure 6.1: The three lowest energy graphene grain boundaries for the $\theta = 30°$ armchair/zigzag interface.

Performing an *ab initio* global optimisation over the configurations at a grain boundary can be expensive. There are two parts to this cost. 1) The cost of the local relaxations, which of course are more taxing then using empirical potentials; and more significantly 2) the exponential scaling of minima as a function of the number degrees of freedom. In this work we present a method of reducing the number of degrees of freedom by inserting chemical knowledge to generate more sensible initial structures which simultaneously reduces the number of optimisation steps required to reach the local minima. The result is a roughly 10 speedup in finding the lowest energy graphene interface configurations compared to Schusteritsch and Pickard.

6.2 Method

6.2.1 Voronoi based boundary generation

Random structure prediction on larger systems (more than 10 particles) owes a big part of its success to the use of sensible constraints to generate initial structures in such a way that excludes high energy parts of the potential energy surface. A light touch is key here, sticking to constraints that apply to *all* low energy structures such as imposing minimum interparticle separations or a small set of randomly chosen symmetry operations. In the case of 2D materials we know that they will adopt a roughly planar arrangement made up of convex polygons. Following the steps in

fig. 6.2 we can generate such structure directly as follows:

1. Populate the plane with a set of random points and generate the Voronoi decomposition (a partition of the plane around seed points such that the cell surrounding each seed contains all points closer to it than to any other seed).
2. Put springs of equal constant along the edges of the Voronoi cells and relax for a small number of iterations.
3. Place atoms at the vertices of the Voronoi cells.
4. (optional) Apply minimum separation algorithm to separate any atoms that remain close.

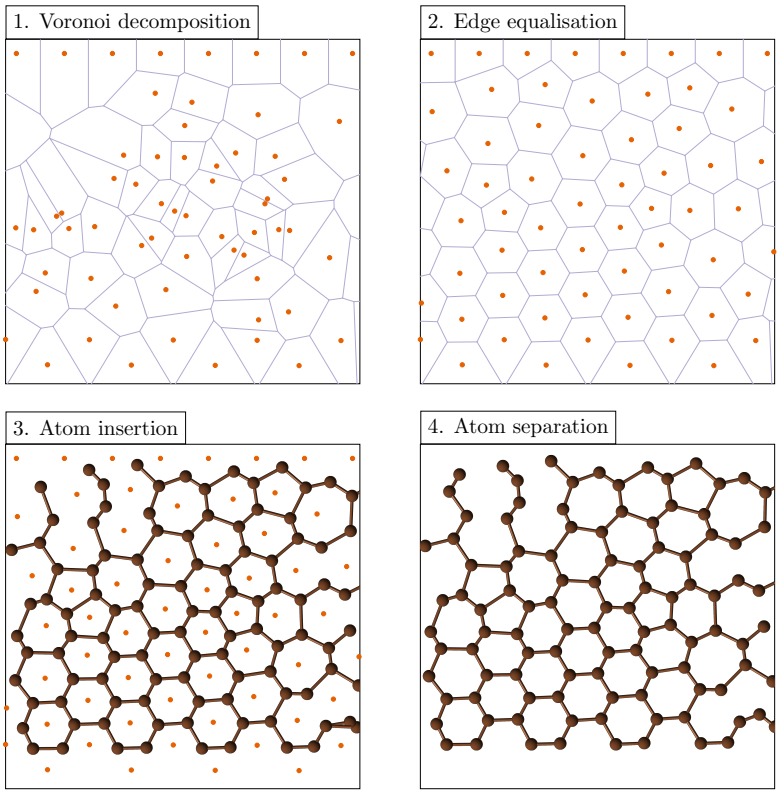


Figure 6.2: The steps for generating a defective graphene sheet using Voronoi decomposition (see main text for description).

Step 2. is crucial as it takes the net and stretches it out bringing the bonds to be roughly similar in length. Already we can see in step 4 that the generated structure is a reasonable approximation to a defective graphene sheet without having done a single step of DFT relaxation. This procedure is similar to that suggested by Nutt and Weller (2009) for generating defective graphene sheets however they use a rather complicated set of heuristics coupled with a Monte Carlo rejection scheme to accomplish essentially what our step 2 does.

The theoretical reduction in the number of degrees of freedom of this method can be quantified as follows. For a perfect graphene sheet each atom is shared between three incident hexagons. Therefore each hexagon can be assigned $\frac{1}{3} \times 6 = 2$ atoms which under our scheme are fully defined by a single point in the plane. Thus, the number of degrees of freedom is halved.

6.2.2 Searching configuration

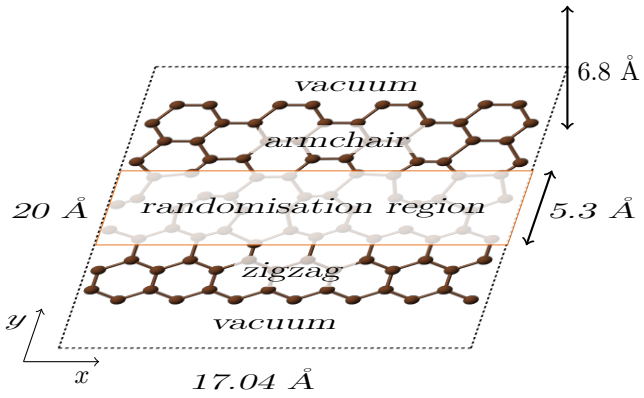


Figure 6.3: The configuration used for generating random interface boundaries using our method.

To generate the random interfaces we adopt the procedure outline by Z.-L. Li et al. and subsequently used by Schusteritsch and Pickard whereby a randomisation region is sandwiched between regions of perfect armchair and zigzag along the centre of a nanoribbon in a periodic unit cell as shown in fig. 6.3. The dangling carbons at top and bottom are terminated with two hydrogen atoms each. Finally after relaxation of the ribbon an xy-periodic grain boundary is created by removing the hydrogen atoms and vacuum and mirroring the ribbon along the y direction creating two grain boundaries separated by several layers of ideal armchair or zigzag.

Initially we search in the same size unit cell as Schusteritsch and Pickard known as (7,0)|(4,4) which has a lattice mismatch of approximately 1% between perfect zigzag/armchair. During searching the unit cell size in the z-direction was 6.8 Å which was increased to 25 Å to calculate final energies for the best few configurations. The quoted interface energies for the grain boundaries are given by

$$\sigma = \frac{1}{2L}(G_{\text{tot}} - n_{\text{C}}\mu_{\text{C}}), \tag{6.1}$$

where G_{tot} is the Gibbs free energy, n_{C} is the number of carbon atoms, L is the unit cell length along the interface and μ_{C} is the chemical potential calculated from relaxed (bulk) graphene.

After initial searching, the unit cell was doubled in width to look for grain boundaries with longer periodicities. The randomisation region was populated with between 13-17 points (each of which corresponds to the centre of a polygon), and twice that for the larger cell, to sample different numbers of atoms in the randomisation region.

6.2.3 Electronic structure calculations

All local minimisations were carried out with the plane-wave pseudopotential code CASTEP using the PBE exchange-correlation functional. Initial searching was carried out with a plane-wave cutoff of 280 eV which was increased to 400 eV for final energy calculations. All calculations were carried out at the Γ point. With these settings, energies were found to be converged to 0.05 meV/atom.

The unit cell was fixed during initial searching and then allowed to relax along the x and y directions when calculating final energies.

6.3 Results

Calculations using a cell width of 17.04 Å produced the same lowest energy grain boundaries, GBs I-III, as found earlier by Schusteritsch and Pickard and others in the case of GBs I and II, alongside higher energy configurations. This provides reassurance that the generation procedure is not inadvertently excluding at least these and other low energy interfaces. Schusteritsch and Pickard report that they relaxed 300 initial configurations to be confident of having found the lowest energy interfaces. In total we carried out 51 relaxations, finding GBs I-III several times after 30. Because both methods involve stochastic sampling this is by no means a rigorous determination of the relative performance but merely qualitative indication. It is also interesting to look at the spread of final energies after relaxation. Schusteritsch and Pickard have a spread of 31.47 eV between the best and worst nanoribbon from an overall set of 700 relaxations, with a spread of 7.52 eV for the top 50. With our method we find an *overall* spread of 6.04 eV. This supports our claim that the algorithm is excluding higher energy portions of the energy landscape by generating more "sensible" initial structures.

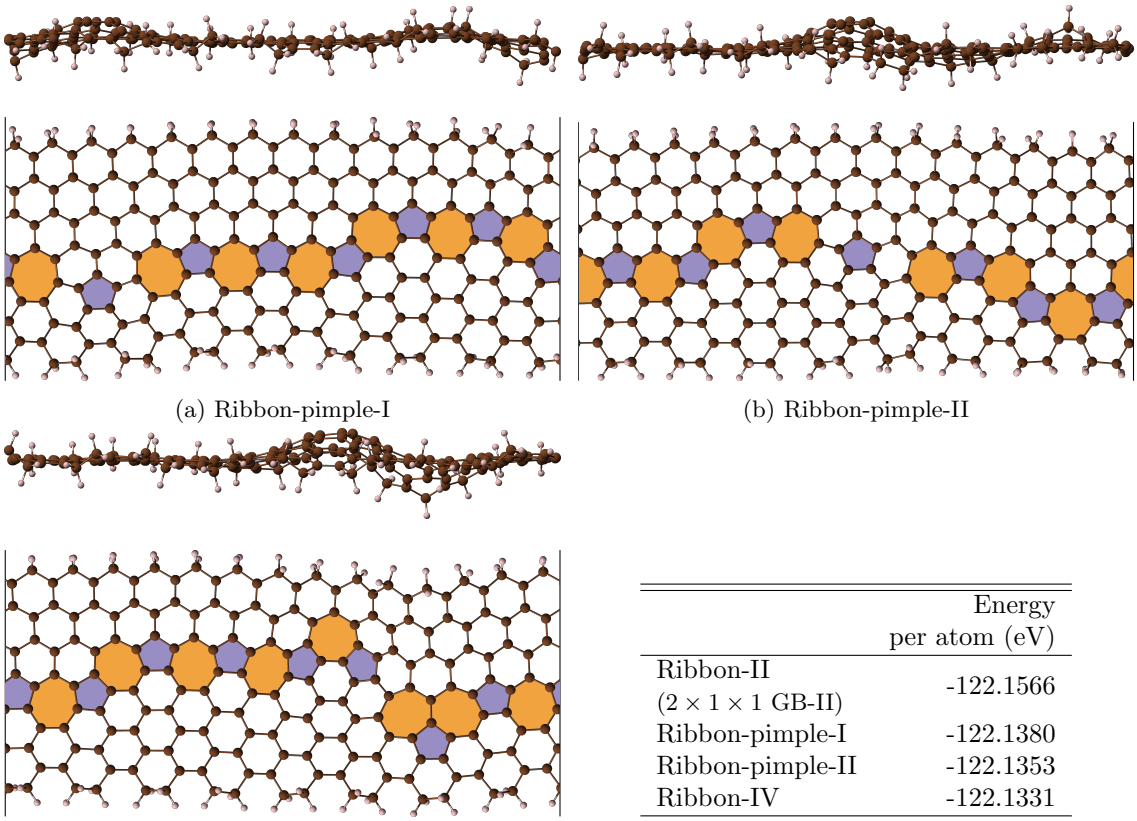
Next we turn our attention to the interface of twice the width. In what follows we present results from 100 relaxations, likely not enough to converge upon the final ground state but enough to demonstrate the potential of the method in large systems.

6.3.1 Nanoribbons

The lowest energy nanoribbon found, ribbon-II, was that of the GB-II supercell repeated twice in the direction of the interface. Figure 6.4 shows the next three lowest along with their energies per atom. In these larger systems a new feature appears along the interface, that of a pentagon surrounded by hexagons which we call a pimple. This naturally introduces stress into the sheet which is relieved by an out of plane buckling around the pimple. Ribbon-pimple-I in particular is very close to the lowest energy, ribbon-II, being just 7.6 meV per atom higher in energy which is just outside of our convergence criteria. Equally the energies of the next two interfaces are not far off although we suspect that more searching may well produce more structures in this energy range.

6.3.2 Grain boundaries

Next we take the previously described ribbons and create periodic grain boundaries that are allowed to relax. The results are show in fig. 6.5. The new interface structures are now nearly perfectly flat



(a) Ribbon-pimple-I

(b) Ribbon-pimple-II

(c) Ribbon-IV

	Energy per atom (eV)
Ribbon-II ($2 \times 1 \times 1$ GB-II)	-122.1566
Ribbon-pimple-I	-122.1380
Ribbon-pimple-II	-122.1353
Ribbon-IV	-122.1331

Energies

Figure 6.4: Graphene nanoribbons showing the interface and side view showing buckling.

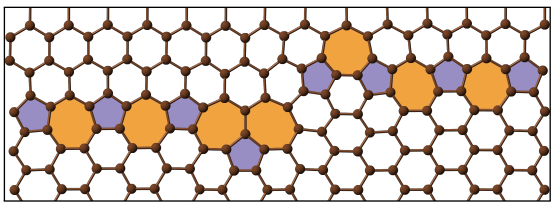
owing to the imposed periodicity. This incurs an energy cost which brings the pimple interfaces to be higher in energy than the previously known interfaces.

6.4 Discussion

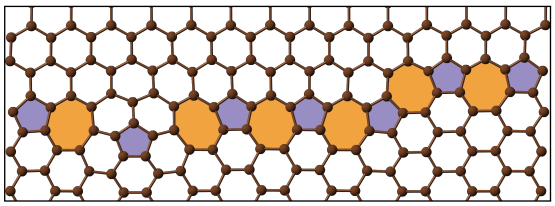
Various experimental studies have reported sinuous grain boundaries in CVD grown graphene (An et al., 2011; Duong et al., 2012; P. Y. Huang et al., 2011; Kim et al., 2011; Yu et al., 2011). This phenomenon was recently the focus of a theoretical study by Z. Zhang et al. (2015) who suggest that this arrangement arises as a way to relieve in plane stress, reducing the overall energy more than the cost of a longer interface. They use a series of building blocks (short sections of interface) to construct various sinuous-shaped boundaries which are then relaxed in atomic simulations. The unit cell sizes used are fairly small with a maximum of 12 polygons along the interface, with our method it would be possible to do a predictive study on such systems with larger unit cells without having to rely on choosing specific building blocks. Another possible avenue for future work is to apply the method to other 2D systems where much less is known about the structure of grain boundaries such as boron nitride and the chalcogenides.

In terms of the method there remains an important unexplored aspect that could lead to a further significant speedup[1]. Our testing shows that the network of C-C bonds generally remains

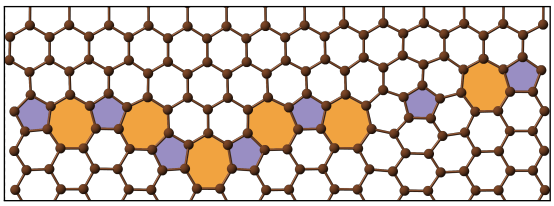
[1]Many thanks to Mike Gillan and particularly Gábor Csányi for crystallising these ideas in a discussion during



(a) GB-IV



(b) GB-pimple



(c) GB-pimple-II

	σ (eV/nm)
GB-I	4.19
GB-II	3.17
GB-III	3.16
GB-IV	4.89
GB-pimple-I	4.55
GB-pimple-II	4.83

Interface energies

Figure 6.5: New graphene grain boundaries found in larger unit cells than considered previously.

the same before and after DFT relaxation. This could be exploited to cull duplicate structures before DFT is even performed. For example, a string that describes the grain boundary could be defined. An example being the degree of the sequence of polygons along the boundary which for fig. 6.1b would be "75757575". This can then be used not only to cull duplicates but also to generate new possible interfaces. It is also the case that particular sequences such as "757" are found to be particularly energetically favourable and these can be used as building blocks rather than individual polygons. With such an approach this method may well be suitable for use with significantly larger cell sizes.

The Voronoi based scheme behind this method could very likely be extended to 3D systems. In this case the seed points would define the centres of coordination polyhedra. Here, the reduction in the number of degrees of freedom would be even greater owing to the greater number of atoms associated with each polyhedron. Indeed Voronoi based methods have been widely used to analyse structures (Aurenhammer, 1991; Damasceno, Engel and Glotzer, 2012; Oda, 2012) but to the best of our knowledge never as a means of automated structure prediction.

6.5 Conclusion

In this chapter we have outlined a new method for predicting the grain boundaries of 2D materials that appreciably outperforms the current state of the art random structure prediction enabling larger system to be tackled. The key step was to introduce chemical knowledge into the procedure for generating initial structures. It is not inconceivable that machine learning or genetic algorithm based prediction schemes could "learn" or encode this knowledge without any human intervention but in the opinion of this author that is to miss the point of these powerful tools. Machine learning and GAs are best applied to those things which we do not or can not know, all prior knowledge,

my viva examination.

assuming it is general enough, should be utilised to its fullest.

7 | Binary Lennard-Jones crystals

If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is . . . that all things are made of atoms . . . attracting each other when they are a little distance apart, but repelling upon being squeezed into one another. In that one sentence, you will see, there is an enormous amount of information about the world, if just a little imagination and thinking are applied.

Richard Feynman

7.1 Introduction

Chapter 3 detailed the history of attempts to elucidate the dominant structure determining factors for a broad range of crystal systems. The trajectory of this field is clearly towards sophisticated analysis of large structure datasets to extract a common set of rules that are as general as possible. Perhaps the most advanced example of this is Villars' recent work (Kong, Luo et al., 2012; Kong, Villars et al., 2012) on determining "materials genes", represented by chemical and bonding parameters, by using information entropy to assess their relative contribution to structure stability.

In this chapter we explore one avenue of within this trajectory by combining high-throughput structure prediction with modern big-data methods to carry out a systematic study of the structures of the binary Lennard-Jones system. Lennard-Jones is used simply as a convenient proxy to describe the features common to almost all bulk forming systems: Long range attraction and short range repulsion. In this sense varying the potential parameters of interaction strength and particle size can be thought of as sweeping through a simple continuous periodic table. Given just these 'knobs' and stoichiometry we ask: What are *all* the crystal structures that can form? The structures found, known and unknown, map out part of the space of possible 3D periodic tilings that is particularly relevant to crystallography and in this sense we think of it as building up an alphabet of structures

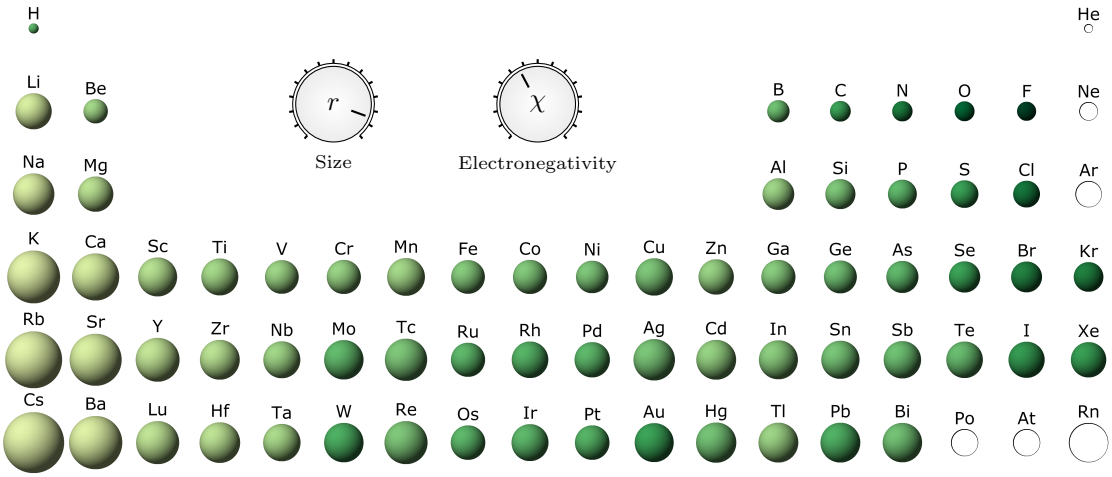


Figure 7.1: The periodic table with spheres drawn according to covalent radii and coloured according to Pauling electronegativities. In nature these two 'knobs' are discreet but using an empirical potential gives us access to a continuous range.

in a similar vain to the mapping out of possible polyhedra in mathematics.

Some applications of this work will be discussed. In self assembly it is often necessary to know the interaction potential that constituents should adopt in order to assemble into a particular target structure. Structure prediction using ionic substitution can also benefit from our large database of structure prototypes to be used as initial seeds. Lastly, the sheer number of structures generated and analysed makes our framework an ideal testbed for the development of high-throughput methods and tools to enable materials discovery.

7.2 Binary mixtures

Binary mixtures of spherical particles are an important reference system for studying the phases of condensed matter. Over the last 80 years such systems have been used extensively to model glasses (Fernández and Harrowell, 2003; Kob and H. C. Anderson, 1995; Lennard-Jones and Devonshire, 1937), colloids (P. Bartlett, 1990), alloys and the freezing of liquid mixtures (Baus, 1987; Haymet, 1987) using a variety of methods including Monte Carlo simulations (Kranendonk and Frenkel, 1989, 1991a,b), cell theory (Cottin and Monson, 1995, 1996), density functional theory (H. Xu and Baus, 1992) and others. Knowledge of the crystalline phases adopted by such systems is important, not only to study the crystallisation process but also in fields such as self-assembly, where the goal is to design constituents such that they spontaneously assemble into a particular target structure.

7.2.1 Hard spheres

Close-packed arrangements of hard spheres represent perhaps the simplest crystal forming system and yet their binary mixtures show a surprising structural richness. Parthé (1961) conducted early work in this field using similar arguments to Laves (1955) (see chapter 3) to rationalise structure in

terms of space packing. In his work Parthé constructed a series of space filling curves as functions of the particle size ratio for various AB and AB_2 structure types. The shape of the curves was found to be well correlated with structure type, providing a representation agnostic structure fingerprint. This was used to identify related and identical structure types. In addition the space-filling values were used to predict possible sequences of structure transitions under the application of pressure. In the absence of interactions, packing fraction is often a good proxy for the free energy and packing arguments have been used to explain complex superlattice structures in experiments on binary mixtures of colloids, nanoparticles and block copolymer micelles some of which are detailed in table 7.1. The argument used is that structures with a high close-packed density have a larger per particle free volume at lower density, this results in a higher entropic contribution and hence lower free energy.

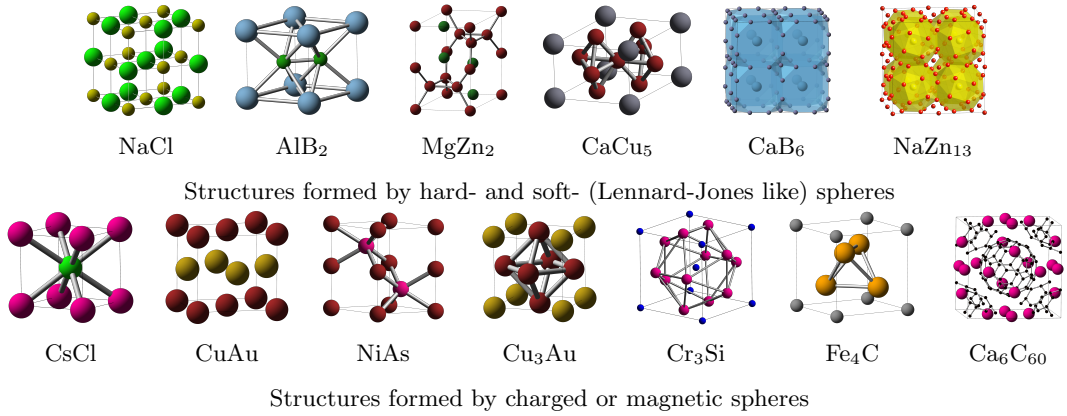


Figure 7.2: Some structures found experimentally in hard-sphere and interacting systems. See table 7.1 for details.

Table 7.1: Details of some experiments on binary mixtures of hard and interacting spheres. The notation L and S is used to label large and small particles respectively. Structures in brackets could not be definitively identified but show likely candidates.

	Reference	Constituents	Particle ratio(s)	Structure(s) formed
Hard sphere or Lennard-Jones like interactions	Sanders (1980)	Natural gem opal (SiO_2)	0.58	AlB_2, $NaZn_{13}$
	P. Bartlett, Ottewill and Pusey (1990, 1992)	Colloid suspension	0.58, 0.62	LS_2 (P6/mmm), LS_{13}
	Hunt, Jardine and P. Bartlett (2000)	Colloidal suspension	$0.39 \rightarrow 0.72$	(NaCl or NiAs), AlB_2, $NaZn_{13}$
	Shevchenko, Dmitri V. Talapin et al. (2002)	$CoPt_3$ nanocrystals	0.58	$CaCu_5$
	Velikov et al. (2002)	Colloidal layers	$0.48 \rightarrow 0.54$	(NaCl or NiAs), LS_2, LS_3
	Saunders and Korgel (2005)	S = gold NC, L = iron NC	0.38	NaCl
	Schofield, Pusey and Radcliffe (2005)	Colloidal suspension	$0.43 \rightarrow 0.53$	AlB_2, $NaZn_{13}$
	Z. Chen, Moore et al. (2007) and Z. Chen and O'Brien (2008)	CdTe and CdSe quantum dots	$0.57 \rightarrow 0.81$	AlB_2, *cub*-$NaZn_{13}$, *ico*-$NaZn_{13}$, $CaCu_5$, $MgZn_2$
	Abbas and Lodge (2008)	Block copolymer micelles	$0.5 \rightarrow 0.6$	*fcc*, LS_{13} ($Fm\bar{3}c$)
	Ye, J. Chen and Murray (2011)	S = Au NC, L = Fe_3O_4	0.33, 0.41	*bcc*-LS_6, CaB_6
Charged or magnetic spheres	Redl et al. (2003)	L = Fe_2O_3 magnetic nanocrystals, S = PeSe quantum dots	$0.54 \rightarrow 0.58$	AlB_2, $CaCu_5$, $NaZn_{13}$
	Leunissen et al. (2005)	Changed colloidal crystals	0.31, 1	NaCl, CsCl, NiAs, LS_6
	Shevchenko, Dmitri V Talapin et al. (2006)	Charged nanoparticles	$0.43 \rightarrow 0.81$	NaCl, CuAu, LS, AlB_2, $MgZn_2$, $MgNi_2$, Cu_3Au, Fe_4C, $CaCu_5$, CaB_6, $NaZn_{13}$, *cub*-$NaZn_{13}$
	Macfarlane et al. (2011)	NCs with DNA linkers	$0.25 \rightarrow 1$ (particles), $0 \rightarrow 3$ (DNA linkers)	CsCl, AlB_2, Cr_3Si, Cs_6C_{60}

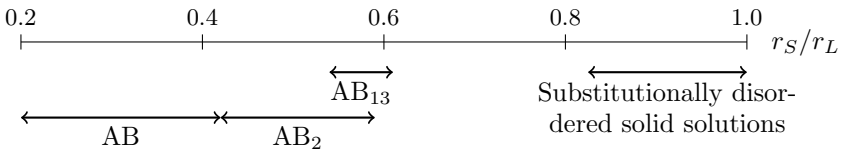


Figure 7.3: Domains of stability for the various hard-sphere stoichiometries as found by Cottin and Monson (1995).

Theoretical progress on hard-sphere systems has often outpaced experiment providing glimpses of the types of structures that can be expected to form. Important early work on the thermodynamic properties of these systems was carried out by Kranendonk and Frenkel (1989, 1991a,b), Denton and Ashcroft (1990) and Cottin and Monson (1995). For example Cottin and Monson used cell theory (Cottin and Monson, 1993) and Monte Carlo simulations to identify the domains of stability, in terms of size ratio, for AB, AB_2 and AB_{13} stoichiometries. Their results are shown in fig. 7.3.

In common with much of the early work (Eldridge, Madden and Frenkel, 1993; Eldridge, Madden, Pusey et al., 1995; Jackson and Ackland, 2007) Cottin and Monson focused on just three stoichiometries and only considered several candidate structures often discarding others based on packing arguments alone. Dissatisfied with this Harrowell and co-workers applied two separate techniques, periodic tiling (Hudson and Harrowell, 2008) and data mining (Kummerfeld, Hudson and Harrowell, 2008), to more systematically study hard sphere structures. Hudson and Harrowell used the Laves tiling method identifying 4 binary crystals with packing fractions better than *fcc* (a requirement for mixing as otherwise each particle type could adopt the *fcc* lattice separately and have a higher packing fraction), two at size ratios not previously considered. In their data mining work Kummerfeld, Hudson and Harrowell used 3000 AB structures from the ICSD, evaluating their packing fraction across the range of particle ratios. Remarkably, the FeB structure was found to exceed or match the packing fraction of all other structures across the full range with only 8 (including NaCl, CsCl and CuTi) matching it at points along the curve.

Filion and Dijkstra (2009) contributed further to the field by making two key advancements. Firstly, they used a genetic algorithm in an attempt to predict new structures. Secondly, they used pressure annealing to squash the resulting structures at constant pressure to calculate the full Gibbs free energies. They report a number of new structures, some with no atomic analogues, not previously considered in sphere packing. For example at $r_S/r_L = 0.82$ they find αIrV, γCuTi, $AuTe_2$ and Ag_2Se to have higher packing fractions than the commonly used Laves phases. However, when free energy is considered only the Laves phases remain stable highlighting a possible pitfall of using packing fraction alone.

Torquato and co-workers also carried out predictive work (Hopkins, Jiao et al., 2011; Hopkins, Stillinger and Torquato, 2012) on binary sphere packings using a novel sphere packing algorithm (Torquato and Jiao, 2010) to explore the composition size-ratio phase diagram. They identified 17

stable phases including seven new structures.

7.2.2 Interacting spheres

Introducing interactions between particles extends the number of accessible structure types and yet obstacles remain to making tunable constituents that assemble into a broad range of phases. Mirkin and co-workers (Macfarlane et al., 2011) have perhaps made the most progress towards this goal using DNA coated gold nanoparticles. As well as nanoparticle size the DNA 'linker' molecule lengths can be varied effectively changing the interaction radius. Macfarlane et al. find that in binary systems both the linker and nanoparticle size ratios determine the assembled structure. Furthermore the sticky ends of the linkers can be varied to, for example, favour or disfavour self interaction making it possible to assemble typical ionic structures such as NaCl.

Theoretical progress has continued apace, mapping out the stable structures for simple potentials over an increasingly broad range of interaction parameters and stoichiometries. Much of this work, including our own, uses the Lennard-Jones potential as a simple model for constituents that attract at long range and repel at short range be they atoms, interacting colloidal particles, polystyrene beads or other spherical particles. The interaction potential for a binary Lennard-Jones system ($i, j \in \{\mathrm{A}, \mathrm{B}\}$) can be written as

$$U_{ij} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right], \tag{7.1}$$

where σ_{ij} act as the length scales (atomic sizes), ϵ_{ij} specify the strength of the interactions and r is the interparticle separation. Without loss of generality it is possible to work in units of the AA interaction (i.e. $\epsilon_{\mathrm{AA}} = \sigma_{\mathrm{AA}} = 1$). This leaves a five dimensional parameter space spanned by $\{\epsilon_{\mathrm{AB}}, \sigma_{\mathrm{AB}}, \epsilon_{\mathrm{BB}}, \sigma_{\mathrm{BB}}, x_{\mathrm{B}}\}$, where x_{B} is the proportion of B particles.

M. J. Vlot et al. (1997) carried out some of the earliest work systematically exploring the binary Lennard-Jones (LJ) parameter space using symmetric particles ($\epsilon_{\mathrm{AA}} = \epsilon_{\mathrm{BB}}$ and $\sigma_{\mathrm{AA}} = \sigma_{\mathrm{BB}}$) at a mole fraction of $x_{\mathrm{B}} = 0.5$. They found that in the range $0.85 \leq \epsilon_{\mathrm{AB}} \leq 2.5$ there is a change from *fcc* to a structure labelled *fcc*(001) (B. M. J. Vlot et al., 1997) around $\epsilon_{\mathrm{AB}} > 2\epsilon_{\mathrm{AA}}$ while demixing occurs for $\epsilon_{\mathrm{AB}} < 1$, perhaps unsurprisingly as in this case AA and BB contacts become energetically favourable. More significantly as σ_{AB} is varied between 0.5 and 1.5 the stable structure type changes from ZnS, to NaCl, to CsCl to *fcc* finally reaching the *fcc*(111) structure confirming the importance of relative particle size found in hard-sphere studies.

Building on this work Fernández and Harrowell (2003, 2004) studied the range $0.6 \leq \sigma_{\mathrm{AB}} \leq 1.1$ at mole fractions of $x_{\mathrm{B}} = 0.25, 0.33$ and 0.5. By exploring these new stoichiometries a host of new structures were found such as the Pd_2Zr, Th_3P_4, Fe_3C structure types and Al_2Cu, originally discussed by Frank and Kasper.

These important studies cast light on selected portions of the binary LJ parameter space and

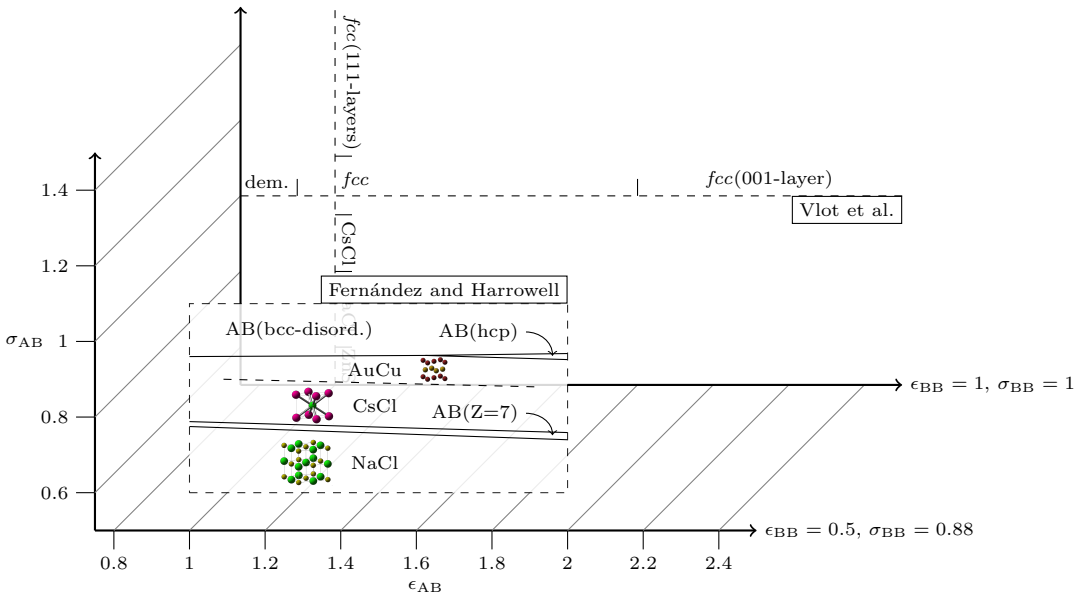


Figure 7.4: Portions of the binary Lennard-Jones ground state structure space that have been mapped out at the AB stoichiometry, Fernández and Harrowell also mapped out AB_2 and AB_3. Much of the space is clearly left unexplored.

helped to explain the stability of certain structure types in terms of interaction strengths and particle radii, however as can be seen in fig. 7.4 much of the parameter space remains unexplored. Furthermore structures may have been missed as much of the earlier work relied on the use of seed structures or coordination polyhedra as building blocks. This work addresses both these limitations by systematically evaluating a much larger portion of parameter space and by using random structure searching to predict stable structures in an unbiased way. Much of this has only recently become possible thanks to the high-throughout capability of modern computers and the maturing of structure prediction methods.

7.3 Big data, high throughput

Invariably a large exploration of the LJ parameter space produces vast quantities of structures and associated data that make analysis nontrivial. To cope with this requires a shift away from the traditional mode of carrying structure calculations to one where structure data is collectively queried and analysed.

7.3.1 Databases

The last decade or so has seen a large growth in the use of structure databases in materials science. Figure 7.5 shows a comparison of the number of entries in a handful of prominent crystal structure databases. Included are the Material Project (Jain et al., 2013), AtomWorks (Yibin Xu, Yamazaki and Villars, 2011), the ICSD (Belsky et al., 2002), The Open Quantum Materials Database (Saal et al., 2013), Pearson's Crystal Structure Database (Villars and Cenzual, 2014), AflowLib (Cur-

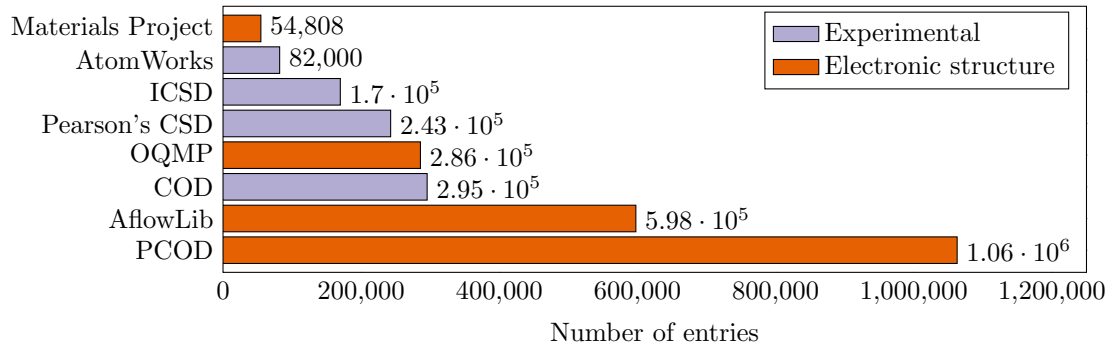


Figure 7.5: A comparison of the number of entries in a range of inorganic crystal structure databases (sampled on the 1st of September 2014).

.

tarolo et al., 2012; Setyawan and Curtarolo, 2010) and the (Predicted) Crystallographic Open Databases (Downs and Hall-Wallace, 2003; Gražulis, Chateigner et al., 2009; Gražulis, Daškevič et al., 2012). Two further examples for which number of entries could not be determined include the American Mineralogist Crystal Structure Database (Downs and Hall-Wallace, 2003) and the Material Properties Open Database (Pepponi, Gražulis and Chateigner, 2012). Most materials databases are either populated from experiment (ICSD, Crystallographic Open Database (COD)) or electronic structure calculations (Materials Project, AflowLib, PCOD, OQMD, ICSD). The real value of such large collections of structure information is in being able to pose new questions, often looking for global or local trends that help to accelerate the discovery process.

In this work we make extensive use of a structure database to store and analyse structure data. As prediction of Lennard-Jones crystal structures is computationally cheap (for small to moderate number of atoms) huge amounts of data can be generated which can then be interrogated. This mode of conducting computational studies can be seen as a shift from that of forming a hypothesis that is then tested by carrying out specific numerical experiments to that where a large number of simulations are carried out to populate the database which is then queried, and crucially visualised, to build up a more complete understanding of relationships between structures and to identify local and global trends which can then be investigated further with additional simulations. In this way of working, then, the structure data takes centre stage with individual calculations becoming practically disposable, emphasis instead being placed on maintaining high throughput.

Figure 7.6 shows a schematic of the use of a structure database in this work. The database is populated from, potentially, a variety of computational resources each performing structure prediction over regions of parameter space. The user interacts directly with the database using a series of tools or for more complex queries custom code. Over the course of this work the database was populated with almost 20 million structures, significantly more than the next largest public structure database. No doubt with increasing interest and use of big data in materials science existing databases will continue to grow and thus our work can be seen as a test bed for where the field is heading.

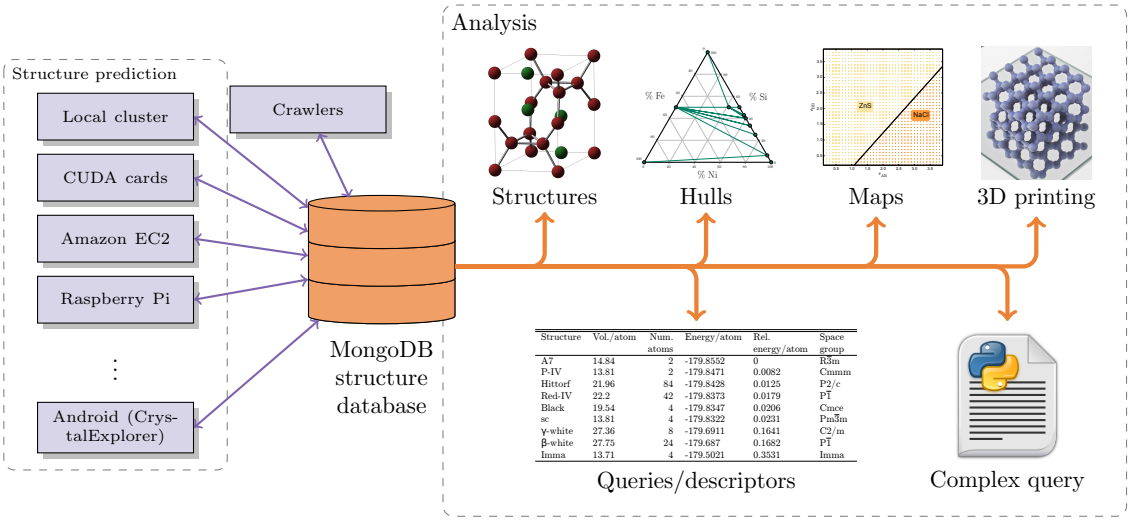


Figure 7.6: High-throughput structure prediction is used to populate a structure database which is then queried by a suite of analysis tools.

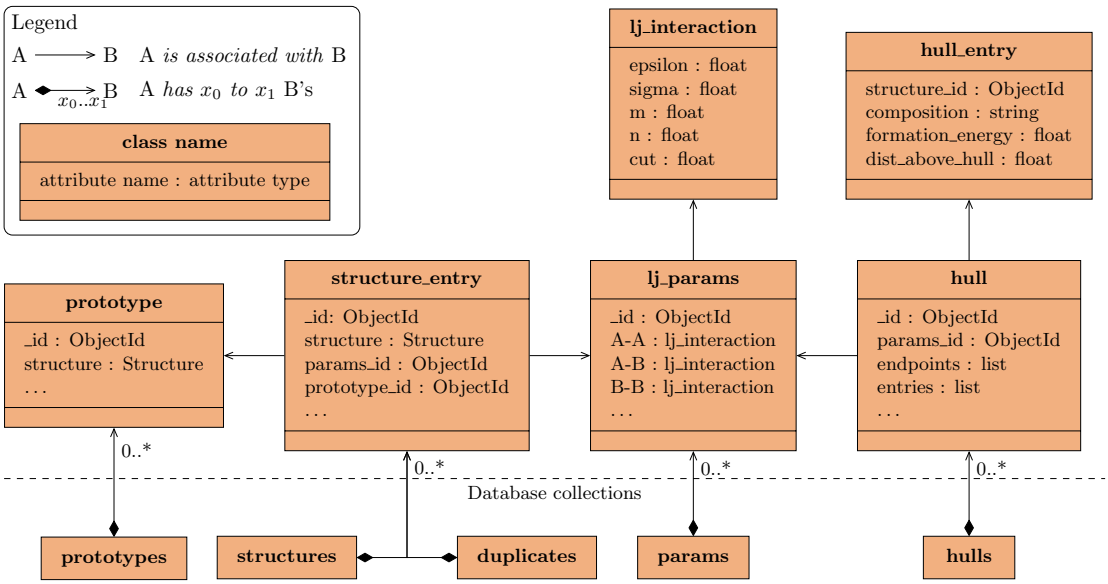


Figure 7.7: A UML class diagram of the relationships between elements in the database.

This work employs the MongoDB database as a back end. MongoDB is a so called schema-less database as it has no fixed structure such as the tables with defined columns found in SQL databases, instead adopting a document-oriented paradigm. *Documents* are dictionaries of key-value pairs where values may themselves be documents. This structure maps well onto objects found in modern object oriented programming languages. Documents are in turn stored within user defined *collections* which form the top level containers in a MongoDB database.

Figure 7.7 shows the database configuration adopted in this work. There are five collections to store structures and their duplicates, structure prototypes, points in parameter space and convex hulls. By default every document has a unique identifier stored in the _id attribute. These are used to form associations between structure entries, parameter points, prototypes and hulls. For example to find the parameter set a structure was calculated with, one can use the params_id from the structure document and perform a search of the **params** collection for the matching _id. As there are many structures for each parameter point this avoids duplicating information which makes searching the database as efficient as possible.

7.3.2 Prototypes

The ability to distinguish between different structure types in an automated way is a key component of this work. In small scale studies this is often done by hand, perhaps using some tools to simplify the procedure. This is not an option for large databases and in fact a large part of the value of a database is derived from accurate labelling of structure types. For example Pearson's Crystal Structure Database is well known for having a well curated set of prototypes. Similarly the ICSD started adding prototypes in 2005 using a procedure described by Allmann and Hinek (2007). We use a modification of this procedure which is based on principles outlined by Lima-de-Faria et al. (1990). We assign a structure to a particular prototype if the two are *isoconfigurational*. According to Lima-de-Faria et al. two structures are said to be *isoconfigurational* if:

i) they are *isopointal*; and

ii) for all corresponding Wyckoff positions, both the crystallographic point configurations (crystallographic orbits) and their geometrical interrelationships are similar.

Two structures to are said be *isopointal* if

i) they have the same space-group type or belong to a pair of enantiomorphic space-group types; and

ii) the atomic positions, occupied either fully or partially at random, are the same in the two structures, i.e. the complete sequence of the occupied Wyckoff positions (including the number of times each Wyckoff position is occupied) is the same for the two structures when the structural data have been standardized.

To check the isopointal condition we use Spglib (Togo, 2015) to first get the structure with a conventional unit cell, this is then standardised using the STRUCTURE TIDY (Parthé, Cenzual and Gladyshevskii, 1993) algorithm to ensure the symmetry settings are comparable. Condition ii) for isoconfigurationality leaves some ambiguity as to the what "similarity of geometrical inter-relationships" means, for this we therefore rely on a similarity comparison algorithm in pymatgen (S. P. Ong et al., 2013) which tries to find a suitable overlap of the two structures to within some tolerance.

In theory it would be possible to use only a structure similarity algorithm to assign prototypes however in practice the computational cost is prohibitive. By storing the Wyckoff sequence of each prototype in the database it is possible to issue a query to retrieve all prototypes that match the isopointal condition which usually produces a small set of structure that can be checked for isoconfigurationality with the structure to be labelled.

7.3.3 High-throughput considerations

The nature of this project, and similarly of big-data electronic structure initiatives, is that there are a huge number of interesting calculations that do not require human selection or intervention to be performed. For example, using a grid spacing of 0.1 and an interval of $[0.1, 3.1]$ for each of the four Lennard-Jones parameters there are $30^4 = 810{,}000$ grid points. Multiplying this by 10 different stoichiometries, and say, 1000 geometry optimisations yields 8.1×10^9. On a large computer it is not unrealistic to perform all these calculation, especially given the pleasingly parallel nature of the problem. However, the results need to be deposited into the database as well as a) validated for errors and spurious results, b) analysed for space group, Pearson symbol, prototype, etc., as structure coordinates alone are rarely useful in a big data context, and ideally c) filtered for duplicates, be added to convex hulls, used for refinement of nearby parameter points, etc. The problem with these operations is that they may easily give rise to bottlenecks which could catastrophically affect overall throughput even if the calculations being carried out are fast. This is particularly true of operations that are non-local to that one structure such as finding prototype, removing duplicates, etc. These operations may need to be carried out eventually but often it is more important to start getting results and have information about the most important subset, in our case the thermodynamically and compositionally stable structures, to allow the investigation to be steered down interesting avenues. By learning the hard way the author has found that one good way to tackle these problems is to consider the entire pipeline that the structure data flows through as one system with many parallel compute nodes and a single central storage in the form of the database. In this model it is now easier to apply standard techniques from parallel computing, many of which are familiar to those in the computational physics community, to analyse and mitigate bottlenecks.

Amdahl's law Amdahl's law (Rodgers, 1985) is commonly used to analyse maximum expected speedup in a parallel system as a function of the number of parallel processing units, n, and the serial fraction of the algorithm or procedure, B. The speedup as a multiple of the serial performance is then

$$S(n) = \frac{1}{B + \frac{1}{n}(1 - B)}.$$

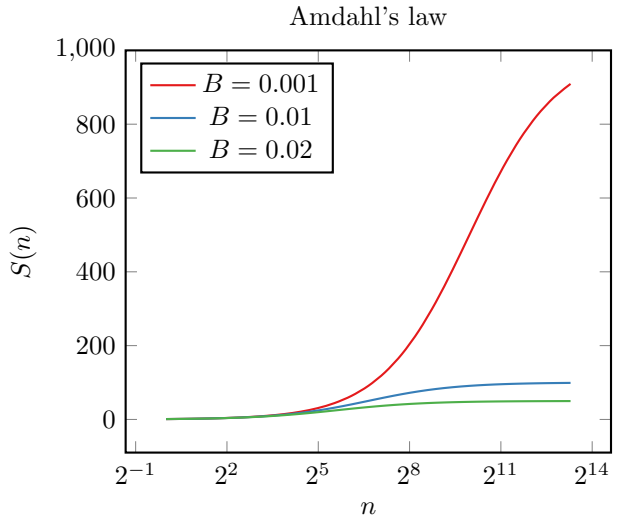


Figure 7.8: Amdahl's law showing the theoretical speedup as a function of the number of parallel processing units, n, for various serial fractions, B. In the case of perfect scaling $S(n) = n$.

The serial fraction in our case comes chiefly, if not solely, from database locking. This may be due to locks by the database software during insertion which are necessary to avoid race conditions or, more disruptively application specific locks that are global. An example of the latter is the insertion of new prototype structures. It is tempting to require that every new structure that is to be inserted into the database is first checked against existing prototypes, if none exists then this structure becomes the prototype and is inserted into the prototypes collection. A Unified Modelling Language (UML) sequence diagram of this procedure is shown in fig. 7.9.

Given that prototypes should be unique, it is necessary to globally lock the **structures** and **prototypes** collections for the duration indicated in red to avoid any new entries being inserted whilst the prototype is determined. Even a very efficient prototyping procedure will result in reaching a speedup plateau in fig. 7.8 given the large n used in this study and therefore this kind of procedure must be avoided at all costs.

One solution to mitigate the aforementioned problem is to decouple the high throughput, data-generation procedures from the analysis and validation. Thus the database becomes a sink for all calculations, the speed of which is now only limited by available resources (network, hard-disks, etc.) which can be scaled according to need. The trade-off comes in the form of having structures in the database that may not yet be validated and analysed, however this cost is often well worth

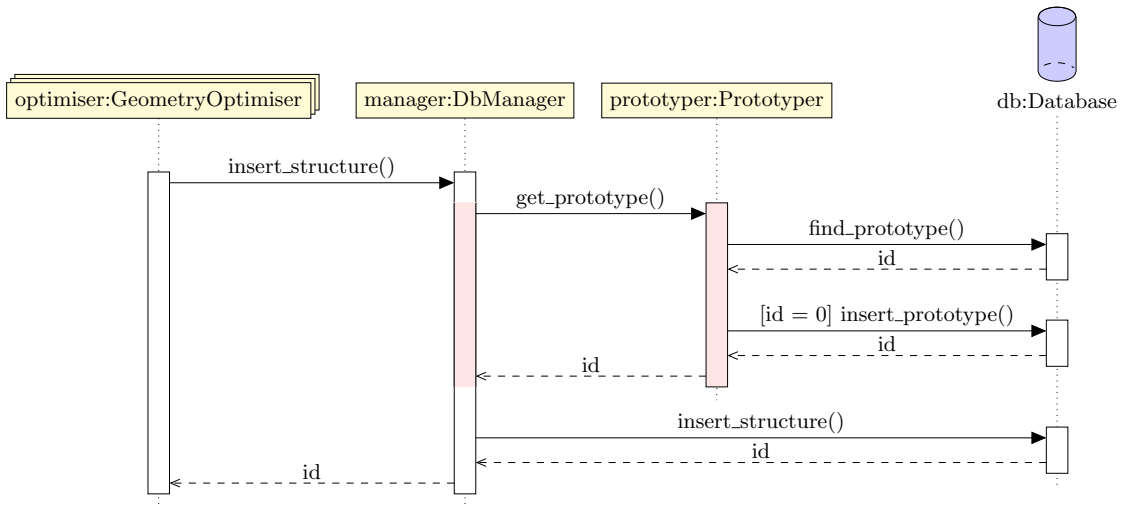


Figure 7.9: UML sequence diagram of a simplified schematic for inserting structures into the database whilst checking for prototypes. Red regions represent global locks during which other callers have to wait.

paying.

Crawlers

In our system validation and nontrivial analysis is handled by crawlers that constantly sweep through parameter points in the database picking up new entries. Multiple crawlers run in parallel always picking the oldest point to be crawled, an atomic procedure whereby the current date is inserted a `last_crawled` field to avoid a race condition. The crawler applies the following functions when it reaches each parameter point:

Prune Remove spurious structures, for example, those with zero energy, huge volumes or unrealistic lattice parameter ratios. Also perform de-duplication.

Prototype Assign prototype to structures, identifying any new prototypes according to the procedure described in section 7.3.2.

Hulls Calculate and store the convex hull.

Refine Gather structures nearby in parameter space an re-optimise to clean up underconverged energy landscapes.

New crawling functions can be trivially added according to need. Where appropriate these functions will act on the N lowest energy structures only as these are commonly most interesting, of course on the next pass the subsequent N will be acted on and so on.

Lazy initialisation

Crawlers allow throughput to remain high and guarantee that analysis and validation will be performed *eventually*, however at times it is necessary to know the answer now. For example,

when generating a structure map it is necessary to know the prototype and compositional stability of the ground state structures. For these cases we employ so called lazy initialisation whereby if the answer is not known (because the crawler has not reached that point) it is computed there and then and stored. This has the disadvantage that the first time a question is posed it may take some time to determine the answer, however we find this to be an acceptable trade-off. The key advantage is that any database contention is localised to just those structures that are involved in the current query (often just ground states) as opposed to all structures at that parameter point.

7.3.4 Structure maps

The use and usefulness of structure maps in materials science has been discussed extensively in chapter 3. In this work we use maps throughout to present data as well as to bring phenomenological trends and features to the fore. To this end we desire an algorithm that is capable of a) capturing bounded regions within parameter space of common type, while b) producing smooth boundaries that are faithful to the underlying data. It seems that no such generic tool is readily available.

Engel and Trebin (2007) created structure maps for a two-dimensional system of monodisperse particles interacting according to the LJ-Gauss potential that are similar to those used in this work, however their boundaries are composed exclusively of vertical or horizontal lines segments giving an undesirable and unfaithful jagged appearance. We developed our own mapping algorithm based on Potrace (Selinger, 2003) that attempts to satisfy the aforementioned criteria. The algorithm takes as its input a set of labelled points in a plane and can be summarised by the following steps:

1. Create a Voronoi diagram of all parameter points (i.e. one Voronoi cell per point).

2. Join cells with a common label that are incident to each other (i.e. share a common point or edge), except in cases where to do so would disconnect two or more incident cells that have different labels.

3. Smooth the resulting boundaries enclosing regions with common labels by

   (a) Checking for possible straight line segments between boundary points that still intersect every edge and choosing the path with the fewest segments.

   (b) Using splines to approximate curves in the path making sure to detect and retain any corners.

Figure fig. 7.10 shows a typical structure map generated using our algorithm. Each dot represents a point in parameter space for which structure prediction was carried out. Dots are coloured according to the colour of the corresponding label. Labels show the structure prototype if known, otherwise the spacegroup is used to differentiate between structures. Grey regions are used to mask off regions that are compositionally unstable.

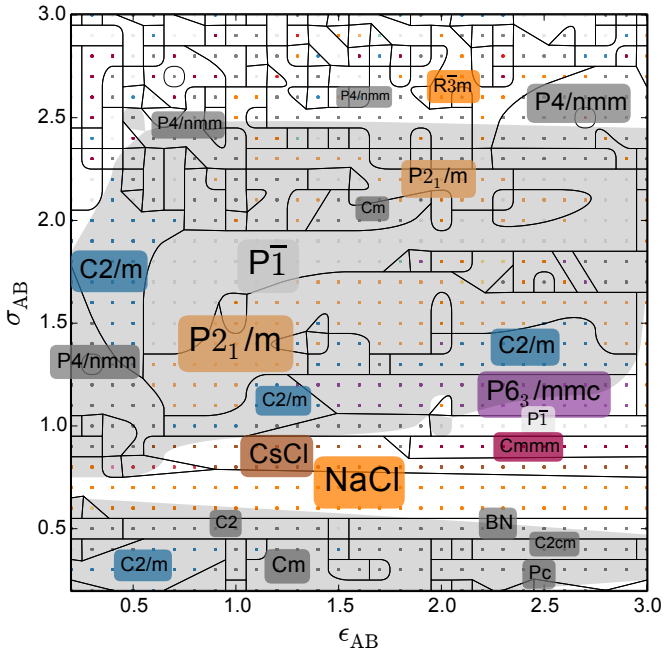


Figure 7.10: A typical structure map used to show bounded regions of common structure type.

7.3.5 Novel architectures

The latter half of the 1990s saw the beginning of a period of relative standardisation of high-performance computing centred around the x86 architecture, thanks largely to its low price per Floating Point Operations Per Second (FLOPS). The last five years or so, however, have seen renewed diversification, again driven by the same financial pressure. In this section we detail some of our findings from investigating a few of these novel architectures.

Generating structures using GPUs

Today, three of the top ten supercomputers in the world use General Purpose Graphics Processing Units (GPGPUs) (TOP500, November 2014) with 50 in the entire list. To examine one possible application of GPGPUs to structure prediction we consider the simple task of constructing a random unit cell with a minimum separation between atoms. This is a commonly applied constraint before any structure relaxation is carried out. An overview of the algorithm used can be found in algorithm 1. Both the random number generator (RNG) and the distance calculation loop can benefit from parallelisation. The main loop was decomposed into blocks of 32^2 iterations (e.g. $i = 1 \rightarrow 32$, $j = 33 \rightarrow 64$), as this allows us to, in the language of GPGPU programming, fill a complete warp. A warp is a set of 32 threads (or 48 for newer cards) that carry out operations in lockstep; therefore by assigning each thread with an atom index i they can check the distance to atom j simultaneously. An additional level of parallelism is available as each GPGPUs has 8 warps that can execute in parallel. Timing results from applying this parallelisation can be found in table 7.2.

Algorithm 1 Pseudocode for generating a set of atoms satisfying a minimum separation constraint using rejection sampling.

```
repeat
    atoms ← GENERATE_RANDOM_ATOMS(num_atoms)
    constraint_satisfied ←True
    for i ← 1, num_atoms do
        for j ← i + 1, num_atoms do
            if DISTANCE(atoms(i), atoms(j))< minsep then
                constraint_satisfied ← False
                break
            end if
        end for
        if ¬constraint_satisfied then
            break
        end if
    end for
until constraint_satisfied
```

The speedup achieved is impressive for a relatively inexpensive consumer-level card, however there are reports of speedups in the region of 100 times serial CPU performance (see e.g. Meel et al. (2008) and Portegies Zwart, Belleman and Geldof (2007)) for some applications. The main bottleneck for this algorithm is that a reduction has to be performed at the end of each block of iterations to communicate the result from each warp to all the others so that they may terminate if the minimum separation condition has been violated. This level of performance penalty can be expected for all algorithms that have a reasonably high ratio of synchronisation to computation; in this case there are approximately 8192 Floating Point Operations (FLOPs) per reduction.

Table 7.2: Speedup results from parallelising constrained random unit cell generation. Timing was carried out using an Intel Xeon E5504 @ 2.0Ghz with a Tesla M1060 GPU.

Code Version	Total (s)	RNG (s)	Compute (s)	Speedup
CPU (serial)	5.58	0.53	5.04	1
GPU with CPU RNG	1.96	0.66	0.266	19
GPU with GPU RNG	1.37	0.0947	0.2365	21.3

Low powered devices

At the opposite end of the spectrum from GPUs lie low power consumption, low cost devices such as the smart phones that have become a ubiquitous part of modern life for many. We have ported the codebase used in this study to the Android platform and released it as a freely available educational app called CrystalExplorer. This was relatively easy thanks to Android support for running native (in this case C++) code, recompiled for the ARM architecture. The performance at the moment is not particularly good, taking roughly 0.1 seconds to perform a geometry relaxation of four atoms on an Nvidia Tegra 3 T33 at 1.2 GHz. A large part of this is down to the fact that most such devices still have a 32-bit architecture. This means it takes several FLOPs even to perform an addition of two double precision numbers. Modern devices are starting to feature 64-bit chips which can be expected to yield a near doubling in performance for scientific applications.

Many such devices are also becoming multicore, with 4 or 8 execution engines becoming common. CrystalExplorer takes advantage of this by performing separate geometry optimisations on each thread.

While not well suited for scientific work, such devices present opportunities for outreach activities. Figure 7.11 shows CrystalExplorer being exhibited at an art show[1]. Participants could play with predicting crystal structures by varying size and interaction strength. This raises the possibility of crowdsourcing calculations from personal devices. Several generated structure were subsequently 3D printed at the UCL Institute of Making and added to the exhibition.

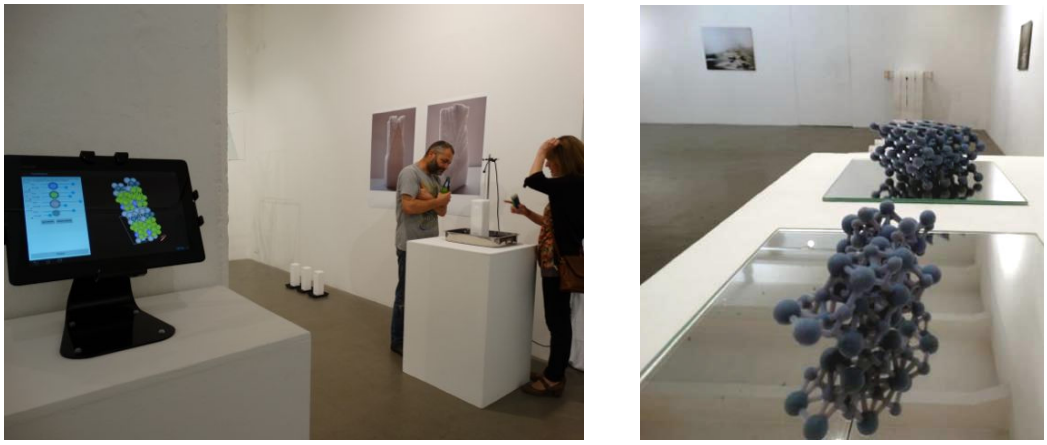


Figure 7.11: CrystalExplorer being exhibited along with 3D printed Lennard-Jones structures generated by visitors on the tablet.

7.3.6 Software infrastructure

The large part of this work involved the development of software to perform structure prediction, automate workflows and read and write results to and from a database. A diagram of the dependencies between major components can be found in fig. 7.12 with a description below. All of the internal components are open source and can be found at https://github.com/muhrin.

SPL

The C++ Structure Prediction Library. This is the central component of the work and includes data structures to manipulate structure and cluster objects, perform geometry optimisation, structure prediction and various forms of analysis including Voronoi partitioning. The code is highly optimised, in particular the routine for calculating distances between points in periodic space is highly optimal outperforming similar Fortran code by a factor of two.

[1]Thanks to Caroline Lambard and Elizabeth Murton for asking me to participate and to UCL StepOut for funding the project.

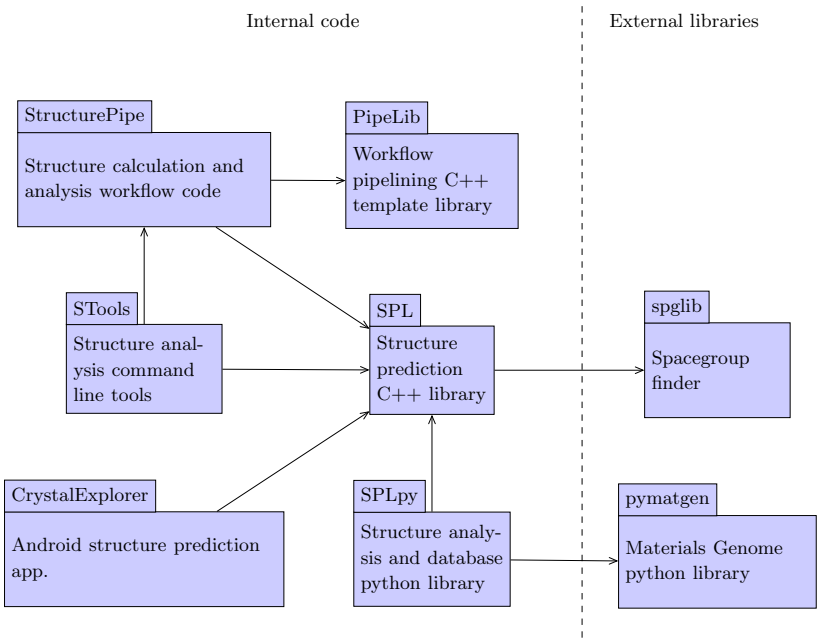


Figure 7.12: The software infrastructure developed as part of this project including some important external dependencies.

SPLpy

A set of python based tools that read output from the SPL and can run StructurePipe to sweep through a range of potential parameters finding the ground state structures. The results are deposited into the previously described structure database. This library also includes database crawlers and an interface to produce structure maps.

PipeLib

A general C++ template library for connecting units of work together in a directed acyclic graph that represents a workflow. The design of this library is motivated by Unix style pipe commands that feed the output of one operation to the next, the major differences being that any data type can be handled, not just text, and that branching is supported if a block supports multiple outputs.

StructurePipe

A C++ application that takes an input file with definitions of what blocks make up a workflow as well as the connections between blocks. Upon running structure objects move through the pipe allowing a flexible sequence of steps to be performed. An example of a common workflow is: Generate random structure → Geometry optimise → Discard duplicates → Determine spacegroup → Write to disk.

STools

A set of command line tools written in C++ to analyse and manipulate structures. The following tools are available.

sconvert Convert between structure file formats.

sdiff Compare two or more structures for similarity using a choice of comparator.

shull Produce a convex hull of the supplied structures.

sinfo Get information about the supplied structure files including energy, pressure, spacegroup, formula, enthalpy of formation, etc. Results can be sorted and printed according to a custom defined input string.

smap Produce a structure map from a text file in the format: `[x] [y] [property]`.

smeasure Measure distances between atoms in a a structure. Can produce ASCII radial distribution output for quick analysis.

CrystalExplorer

An Android app. that uses SPL to perform structure prediction and visualises the ground state found for a chosen set of LJ parameters.

7.4 Method

7.4.1 Determining ground state

To determine the lowest energy structures M. J. Vlot et al. used large systems of 400-1200 particles that were evolved using a Monte Carlo scheme. By taking snapshots at various times they were able to extract candidate structures and calculate their Gibbs free energies. In contrast to this Fernández and Harrowell (2003) used chemical intuition to choose a set of coordination polyhedra that were used to create crystal structure seeds whose energy was locally minimised. Both these approaches are limited in the scope of structures sampled by virtue of the fact that they do not constitute a global optimisation of the underlying energy landscape.

In this study we use random structure searching to find the ground state structures. To ensure the initial structures are 'sensible', an appropriate minimum interparticle distance is used that is given by three fifths of the minimum energy separation of the LJ potential, $r_{ij}^0 = 2^{1/6}\sigma_{ij}$. This is also used to calculate a semi-random initial unit cell volume. To judge when the landscape is well

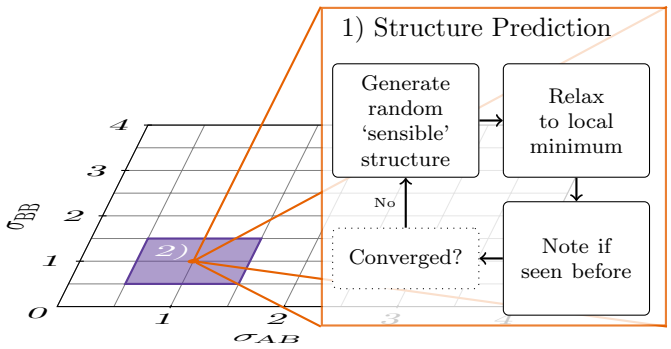


Figure 7.13: Structure prediction over a grid of parameter space values. In step 2) the unique ground state structures from surrounding grid points are re-optimised at the current grid point.

sampled we use the criteria that the lowest energy structure must have appeared many times and no new low lying structure are being found.

As shown in fig. 7.13 a structure prediction is performed over a grid of parameter space values. Once this is complete the unique structures surrounding each grid point are re-optimised with that parameter set to make sure no structures have been missed in the original run.

7.5 Results

Table 7.3: Some statistics about the data produced in this work.

Maximum parameter extents studied	$\sigma_{ij}, \epsilon_{ij}$: $0.1 \rightarrow 4$
No. parameter points	38,923
Total structures	19,089,939
No. prototypes	90,525
Distinct compositions	47

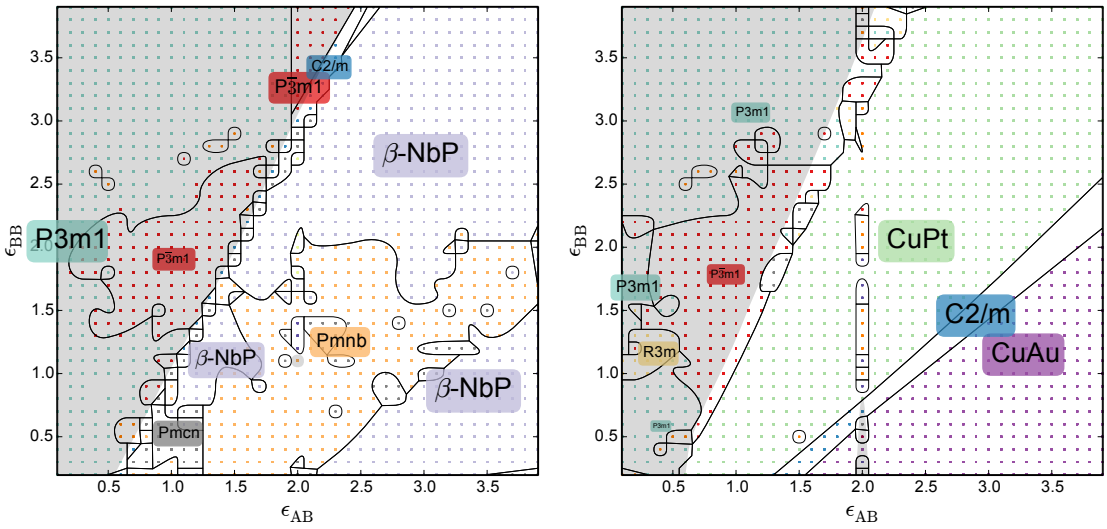
Table 7.3 shows an overview of some of the statistics of the database generated in this work. With such a vast quantity of structures it is worth starting with an orientation of portions of parameter space at fixed stoichiometry followed by results with composition allowed to vary. Finally specific regions of interest are investigated to highlight connections to applications and interesting finds.

7.5.1 Bird's eye view of parameter space

$x_{\mathbf{B}} = 0.5$

We first consider equimolar structures, looking at ϵ_{AB}-ϵ_{BB}, σ_{AB}-σ_{BB} and σ_{AB}-ϵ_{AB} spaces in turn.

$\epsilon_{\mathbf{AB}}$-$\epsilon_{\mathbf{BB}}$ maps Figure 7.14a shows results with all length scales set to 1 and energy scales allowed to vary. This portion of parameter space is fairly simple with only several phases accessible. This can be compared to fig. 1 in M. J. Vlot et al. (1997) (with $e = \epsilon_{\mathrm{AB}}$, $\epsilon_{\mathrm{BB}} = 1$) with condensed results shown in our fig. 7.4. In this case the structure types found differ, possibly as a result of



(a) $\sigma_{AB} = 1$, $\sigma_{BB} = 1$ (b) $\sigma_{AB} = 1.1$, $\sigma_{BB} = 1$

Figure 7.14: Structure maps for the equimolar case (i.e. $x_B = 0.5$) when varying the energy scales.

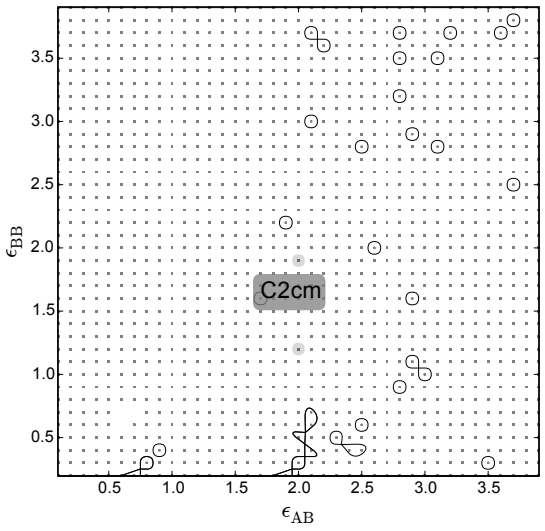
the fact that M. J. Vlot et al. use finite temperature. Increasing σ_{AB} to 1.1 (fig. 7.14b) we see that the region of stability for the P3m1 phase is largely unchanged while $\beta-$NbP is now replaced by three new structures.

Figure 7.15 shows further portions of the ϵ_{AB}-ϵ_{BB} space covering a range of σ_{AB} values with σ_{BB} always fixed to 1. Maps figs. 7.15a and 7.15b show a particularly simple portion of parameter space with only one or two accessible phases. The complexity increases in maps figs. 7.15c and 7.15d covering several close-packed and typical ionic structures including NaCl and CsCl. Finally in maps figs. 7.15e and 7.15f we see regions where the same two or more phases alternate in showing up as the ground state, their energy difference being smaller than numerical precision. This is likely to be an effect of the truncation of the potential after 2.5σ and may be resolved if this were increased.

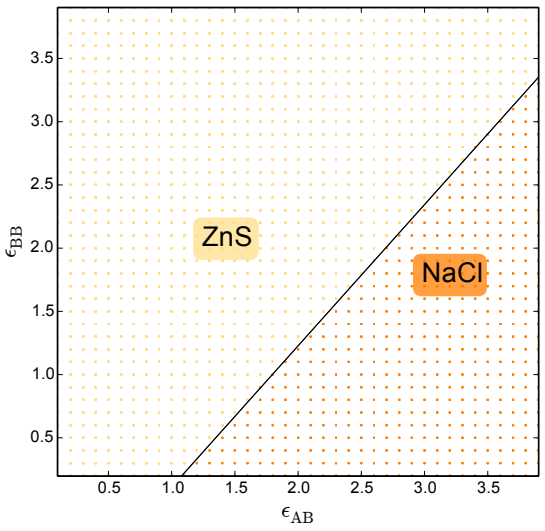
Combining rule The maps shown thus far show a curious feature whereby the boundaries separating phases tend to be approximately straight emanating radially from a common point around (0,-1). This suggests there may be a redundancy in the parameter space which could be eliminated in cases such as this where the ground state structures are of primary interest. This redundancy can be expressed as a combining rule.

A combining rule is an expression for the mixed potential parameters (AB) in terms of the self-interaction parameters (AA and BB). A commonly used example is the Berthelot rule (Berthelot, 1898) given by: $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$. The redundancy in our plots can be seen most clearly by drawing a vector with norm k starting from (0,-1) as in fig. 7.16a. Clearly as k is varied the ground state structure type does not change. This redundancy can be encoded by the following choice of combining rule:

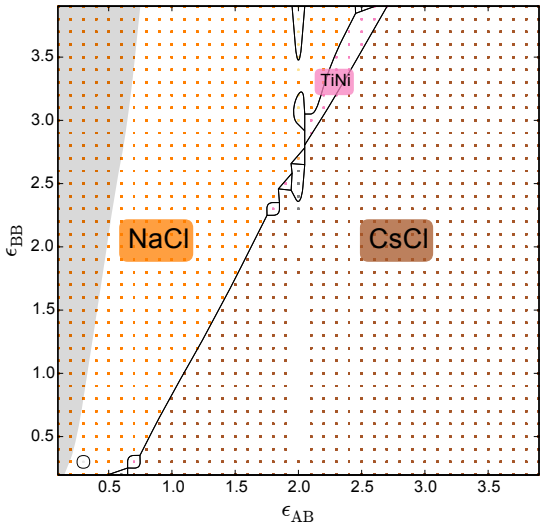
$$\epsilon_{ij} = \sqrt{k^2 - (\epsilon_{ii} + \epsilon_{jj})^2}, \tag{7.2}$$



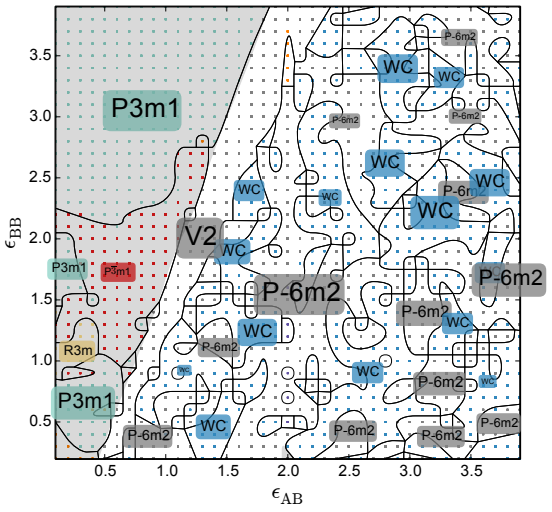
(a) $\sigma_{AB} = 0.4$, $\sigma_{BB} = 1$



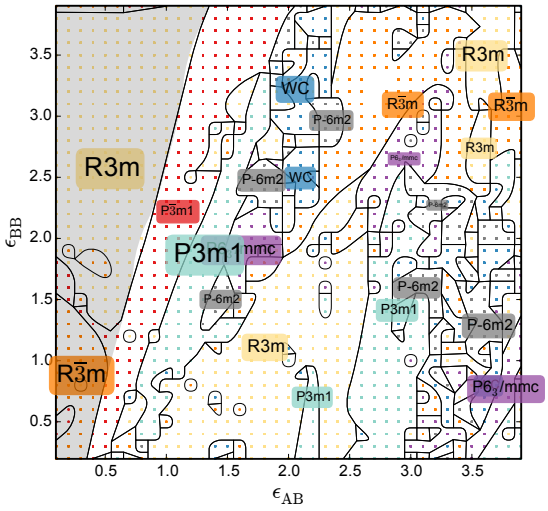
(b) $\sigma_{AB} = 0.6$, $\sigma_{BB} = 1$



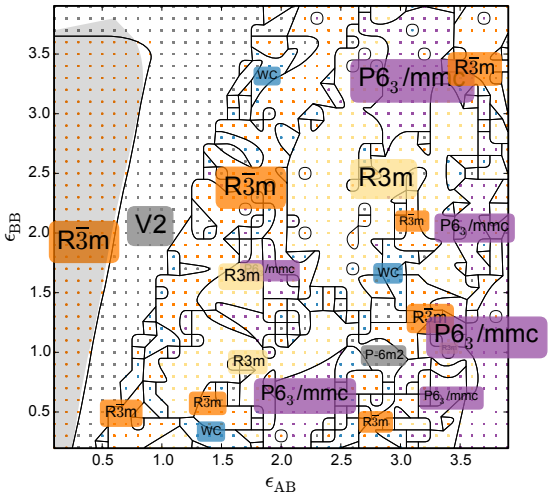
(c) $\sigma_{AB} = 0.8$, $\sigma_{BB} = 1$



(d) $\sigma_{AB} = 1.2$, $\sigma_{BB} = 1$



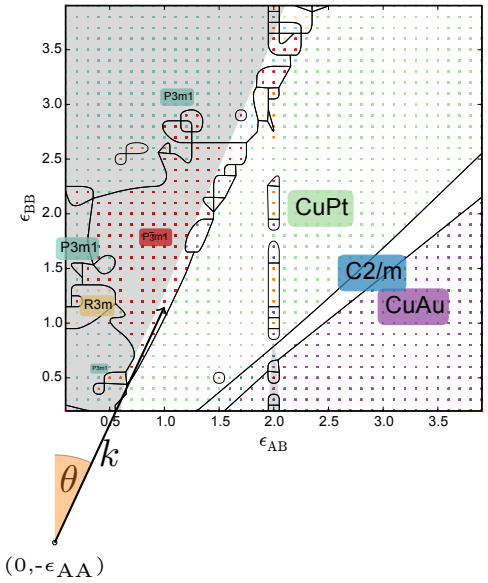
(e) $\sigma_{AB} = 1.4$, $\sigma_{BB} = 1$



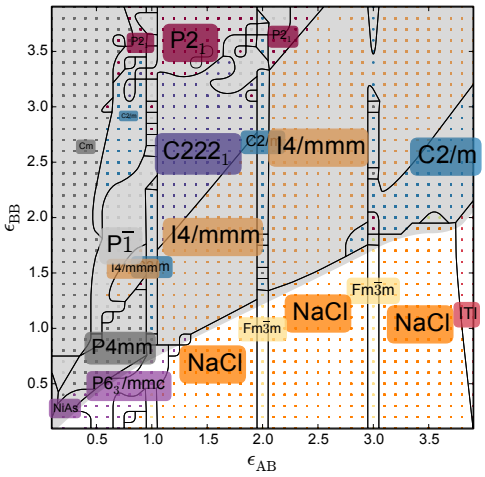
(f) $\sigma_{AB} = 1.6$, $\sigma_{BB} = 1$

Figure 7.15: Portions of the ϵ_{AB}-ϵ_{BB} space at various values of σ_{AB} with $\sigma_{BB} = 1$.

where k is an energy constant and can be chosen to be ϵ_{AA}.



(a) $\sigma_{AB} = 1.1$, $\sigma_{BB} = 1$ showing invariance of ground state structure type with vector length k.



(b) $\sigma_{AB} = 0.7$, $\sigma_{BB} = 0.4 \neq \sigma_{AA}$ showing the breakdown of invariance with variation in k in eq. (7.2).

Figure 7.16: ϵ_{AB}-ϵ_{BB} maps showing regions where our combining rule captures an invariance of ground state structure and where it breaks down.

The invariance in ground state structure holds for particles with $\sigma_{AA} = \sigma_{BB}$ even as σ_{AB} is varied. However, if the self-radii (σ_{ii}) are unequal the invariance no longer holds true. Figure 7.16b shows an example of this with $\sigma_{AB} = 0.7, \sigma_{BB} = 0.4$, where some radial banding remains, however there are regions where varying k will clearly change the ground state structure. Nevertheless our rule captures an intrinsic property of some parts of parameter space not captured by Berthelot, and indeed the adequacy of the Berthelot rule has been questioned (Delhommelle and Millié, 2001). It may be that in certain cases our rule provides a more fitting alternative.

σ_{AB}-σ_{BB} maps Next we vary σ_{AB} and σ_{BB} with ϵ_{ij} fixed to $\epsilon_{AB} = 1.5$, $\epsilon_{BB} = 0.5$, a choice that favours mixed structures. The resulting map is show in fig. 7.17 and shows an incredible richness of structure. It may be compared directly to fig. 3 in Fernández and Harrowell (2004) which is plotted from similar results at constant $\sigma_{BB} = 0.88$ and shows the same series of structures. The range of accessible phases is significantly increased over any of the energy space plots. This may have been expected from the previously discussed importance of space filling to structure stability, however to the best of our knowledge the shear number of phases adopted by the binary Lennard-Jones system has never been highlighted before.

ϵ_{AB}-σ_{AB} maps Figure 7.18 shows ϵ_{AB} varying with σ_{AB} with $\epsilon_{BB} = 0.5$ and $\sigma_{BB} = 0.88$ and can be compared directly with the work of Fernández and Harrowell (2003) shown in fig. 7.4. The horizontal banding further emphasises the significantly greater role played by length rather than energy scales in determining the ground state.

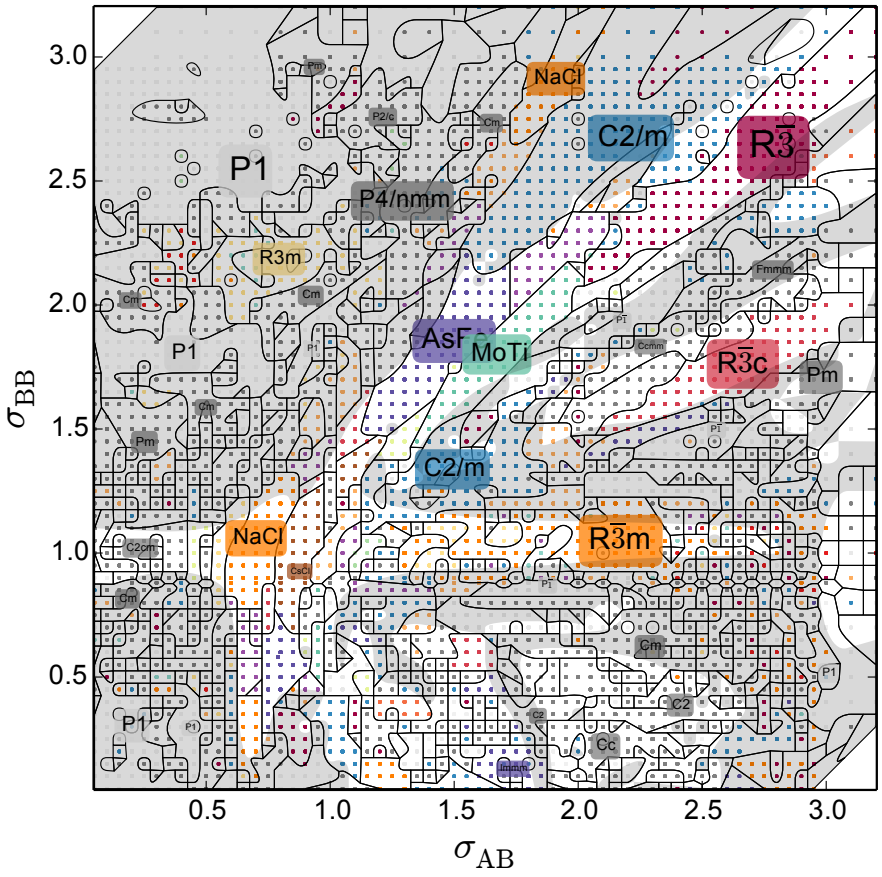


Figure 7.17: $\epsilon_{AB} = 1.5$, $\epsilon_{BB} = 0.5$.

Variable stoichiometry

Next we turn our attention to results for stoichiometries other than AB. In much of the previous theoretical work compositional stability is overlooked, however in almost all cases this information is vital as without it there is no way to know if a given structure will phase segregate.

For each parameter point, structure prediction has been performed for all structures with up to 8 A or B particles. Given the structural richness of the σ_{AB}-σ_{BB} space we continue to focus on this portion of parameter space, again choosing $\epsilon_{AB} = 1.5$ and $\epsilon_{BB} = 0.5$ to favour mixing. In figs. 7.19 and 7.20 we show a series of maps at selected stoichiometries from this region. Immediately striking is just how little of this space is compositionally stable. Indeed certain stoichiometries such as A_4B_5 have no stable points. Many of the other maps show pockets of compositional stability that intersect a variety of structure types rather than having a mask boundary that coincides with that of a particular phase. A number of known structure types feature as well as unknown structures, some of which are analysed in detail below.

Some of the maps show signs of noise where individual parameter points pop into stability within a large region of one phase such as the R$\bar{3}$m region in A_3B. These are likely to simply be a symptom of underconvergence of the landscape at that point.

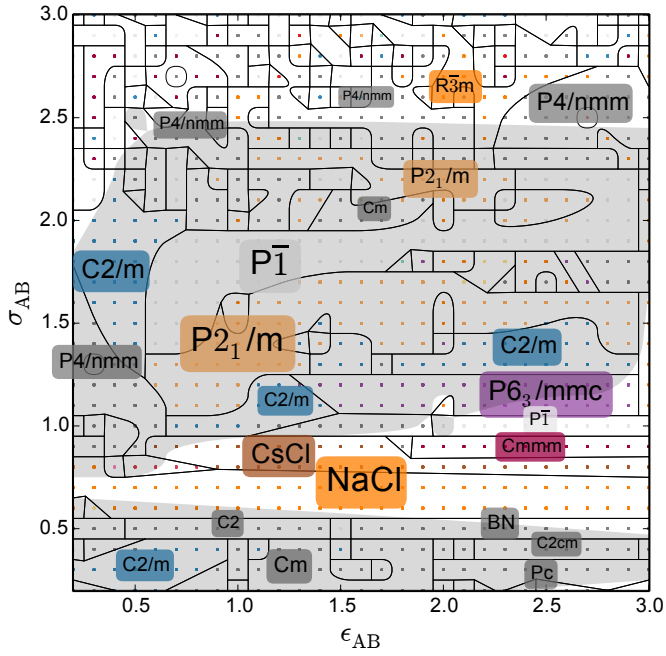


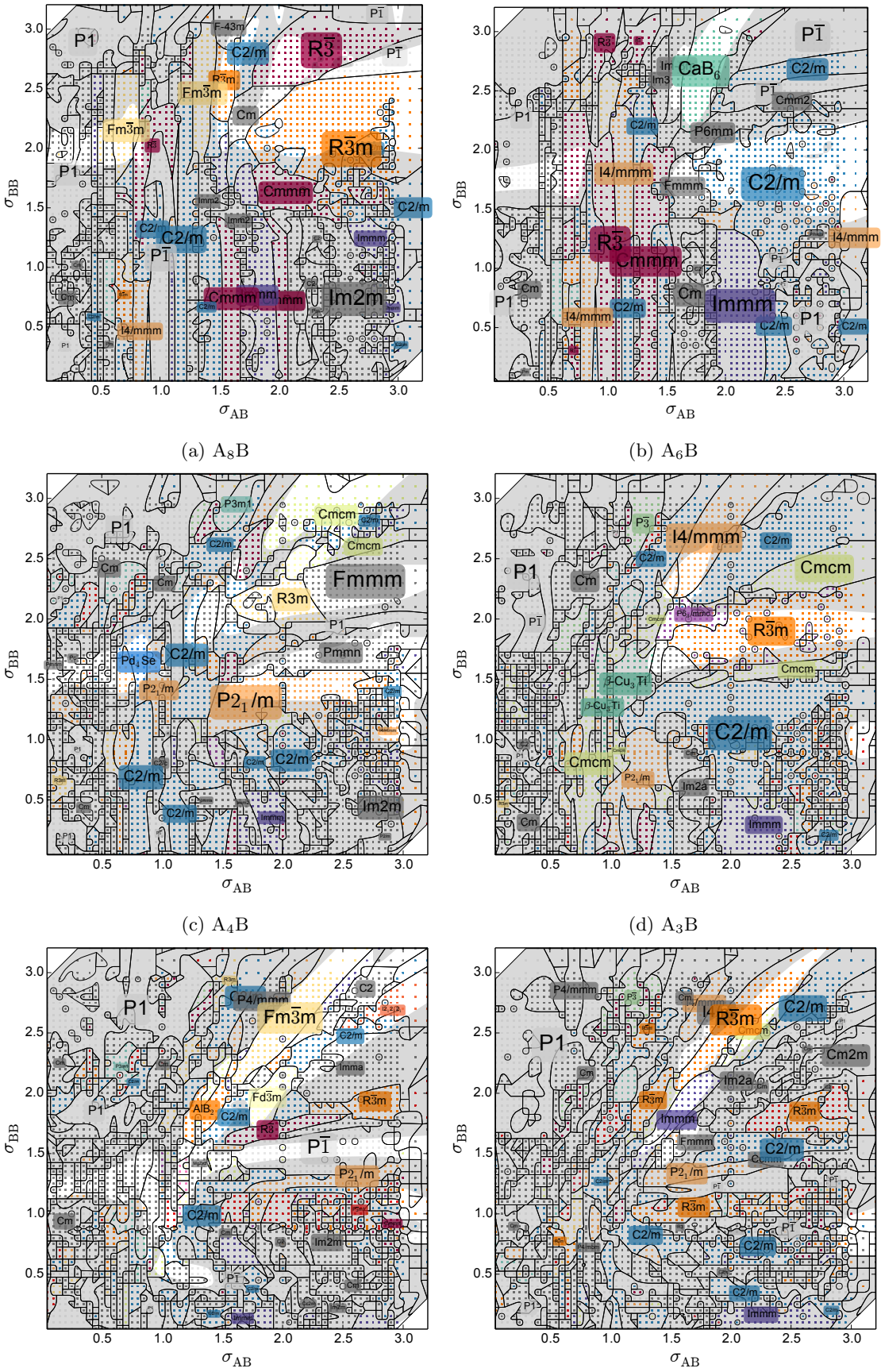
Figure 7.18: ϵ_{AB}-σ_{AB} parameter space with $\epsilon_{BB} = 0.5$ and $\sigma_{BB} = 0.88$. The horizontal banding of structures in this plot confirms the importance of length scales to structure stability.

7.5.2 Up close: Some interesting finds

New structures

In 2007 Hart (2007) wrote a paper asking "Where are nature's missing structures?". In this work Hart uses a combinatorial method (Hart and Forcade, 2008) to generate geometrically simple structures arguing that these are often found at energy landscape extrema. By way of example Hart generates all possible A-B structures using *fcc* building blocks having 4 atoms or less in the unit cell. From the 17 possible structures, 9 have physical manifestations with 8 remaining 'missing' (i.e. unobserved). From these three have been previously predicted while the remainder are described for the first time. Hart goes on to analyse the full set of 17 and suggests reasons for the existence or otherwise of particular structures based on the distributions of bond types (i.e. A-A, A-B, B-B) and neighbour distances by defining a likelihood scale. In table 7.4 we list all the structures ordered by likelihood. The relative likelihood within, for example, the known structures is unlikely to have any real meaning however the scale does seem to roughly order the groups of existing, predicted and not predicted structures in descending order.

One structure in particular, $CdPt_3$, has become an object of fascination in the literature. This structure, first predicted to be stable in the Ag-Pd system by Müller and Zunger (2001) falls between the structures with known manifestations and those without on Hart's likelihood scale. Further first principles calculations (Hart, 2009) suggest that this structure is also stable in the Cd-Pt and Pd-Pt systems. Indeed some of the most recent experimental evidence (Mshumi et al., 2014) weighs in favour of its existence in the Cd-Pt system. In our database this structure is a



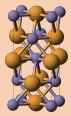
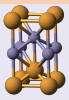
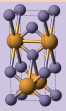
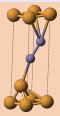
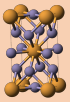
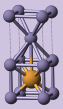
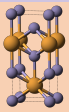
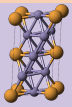
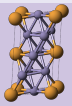
(a) A_8B (b) A_6B

(c) A_4B (d) A_3B

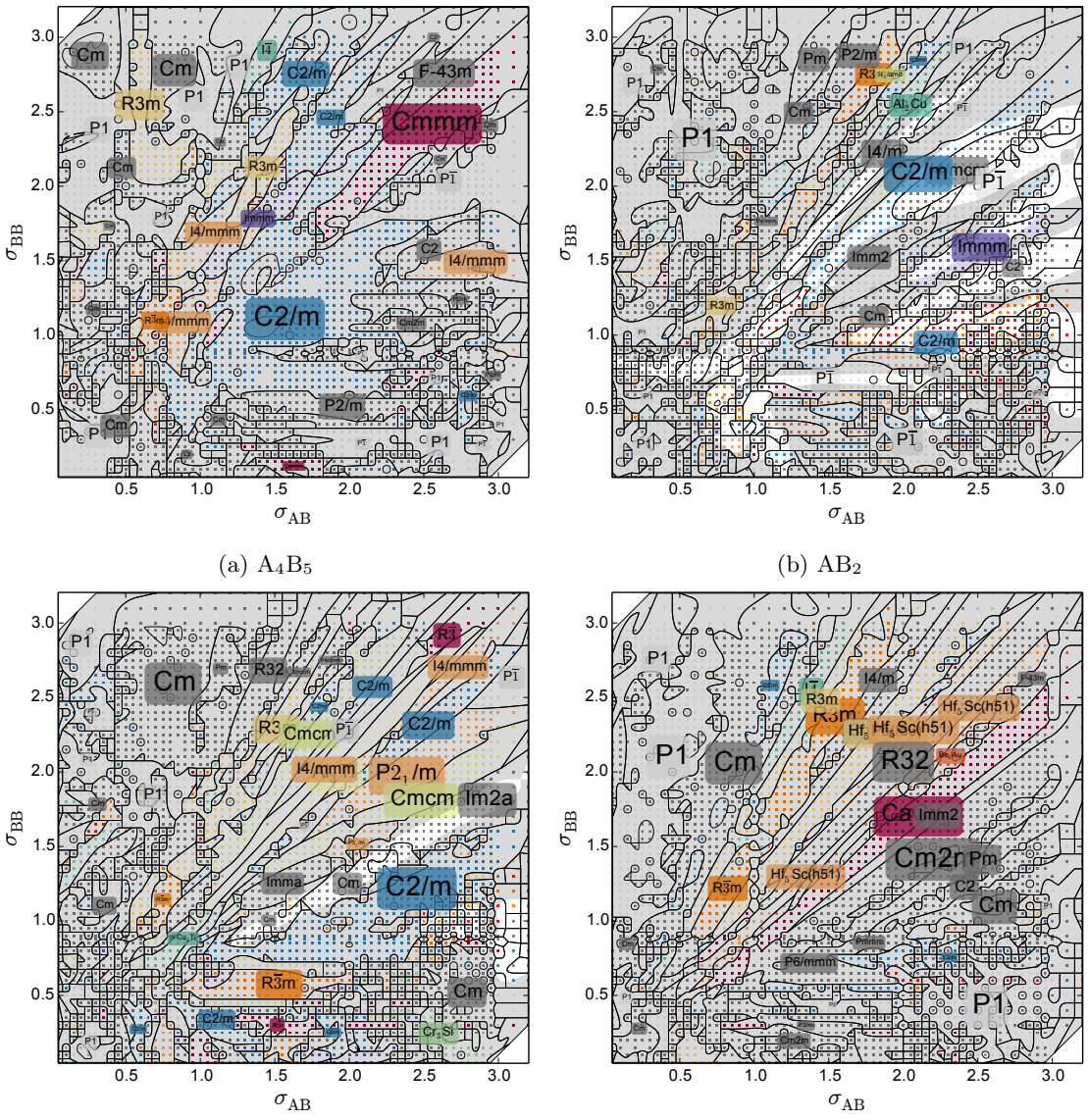
(e) A_2B (f) A_3B_2

Figure 7.19: Selected stoichiometries at $\epsilon_{AB} = 1.5$, $\epsilon_{BB} = 0.5$.

Table 7.4: Structures generated by Hart (2007) ordered according to likelihood to have physical manifestations (top most likely) as defined therein. Compounds in quotes indicate the system that structure is predicted to exist in. Row colours (key below) are based on findings from our database where the condition applies if true at any point in parameter space.

Label	Exists	Predicted	Not predicted
$L1_0$			
(NbP)			
B11			
$L1_2$			
$L1_1$			
$C11_b$			
C6			
AB(111)			
$D0_{22}$			
'AgZr'			
'$CdPt_3$'			
'Pt_3Tc'			
($MoPt_2$)			
$AB_3(110)$			
$AB_3(111)$			
AB(113)			
$AB_3(113)$			

A B Stable Ground state (off hull) Metastable



(a) A_4B_5 (b) AB_2

(c) AB_3 (d) AB_5

Figure 7.20: Selected stoichiometries at $\epsilon_{AB} = 1.5$, $\epsilon_{BB} = 0.5$.

ground state but is not found to be stoichiometrically stable.

We attempt to make some connection between the existence of a structure and its stability in Lennard-Jones. In table 7.4 we see that all structures but one that are stable and on the hull have known physical manifestations, including $MoPt_2$ which comes low down on Hart's scale. The exception of AB(111) also has a high likelihood on Hart's scale. There are three structures with known manifestations which are not found to be stable and on the hull. This may indicate that this criteria is more prone to false-negatives than false-positives. None of the four least likely structures are found to be compositionally stable. Overall in this set of data there is not enough evidence to draw any conclusions and therefore a more extensive analysis will form a part of future work.

These kinds of stories give heart to the theoretician that the days of commonly and accurately predicting structures before experiment are just around the corner. There are also important lessons here for how to make this vision a reality namely combining a wealth of structures with

some way of suggesting existence likelihood. In Hart's case this came in the form of rationalisation to home in on the structure types that are in general likely to exist, followed by direct energy evaluation to pinpoint the physical systems to look for manifestation in.

In table 7.5 we list several structures found in our database that have cubic symmetry and were not found in the ICSD. Here we describe each of these in turn.

η1 This structure exhibits purely A-B bonding with A particles being 6-fold coordinated and B particles 8-fold. Bond distances are bimodal with short (0.99 σ_{AA}) and long (1.145 σ_{AA}) bonds occurring in a 1:1 ratio.

η2 This structure is composed of a series of truncated octahedra surrounding each of the large A particle with vertices populated by small B particles. The polyhedra with As at their centre make up the Kelvin structure proposed as a solution to the most efficient packing of foam bubbles by Thomson (1887). However, we know of no previously reported structure where there are additional particles on the vertices of the polyhedra.

η3 This structure is composed of a series of tiled rhombicuboctahedra surrounding each of the large A particle with vertices populated by small B particles.

η4 In this structure the large A atoms form an *fcc* lattice. The smaller B atoms form a further *fcc* lattice at half the scale and shifted by a quarter lattice vector.

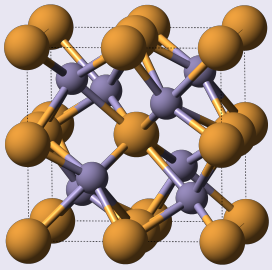
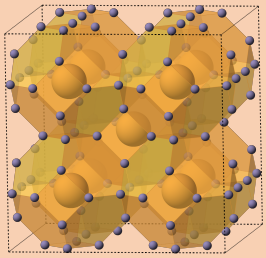
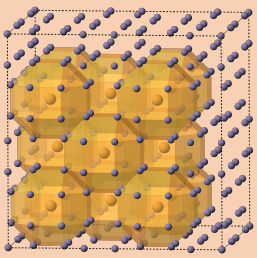
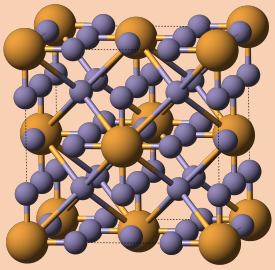
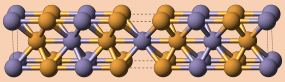
η5 This structure has a central mirror plane of B atoms in a square lattice, to either side are a series of corrugated hexagonal layers in the sequence BABAB. The structure appears in portions of the parameter space where A and B are of a similar size and the AB interaction strength is large relative to the self-interactions.

Laves phases

The Laves phases discussed in chapter 3 while being fundamentally interesting also hold a significant practical interest because many of the materials adopting these structures exhibit a range of useful properties including superconductivity, high hydrogen-absorbing capability and retention of mechanical properties at high temperature. The TCP nature of these structures means there is an ideal size ratio of r_L/r_S of $\sqrt{3/2} \sim 1.225$ and indeed in our maps the cubic $MgCu_2$ is stable near $\sigma_{BB} = \sqrt{3/2}$ (A_2B - fig. 7.19e) and $\sigma_{BB} = \sqrt{2/3}$ (AB_2 - fig. 7.20b). The stability range in σ_{AB} is also as expected for hard-spheres, near $\sigma_{AB} = (\sigma_{AA} + \sigma_{BB})/2$. We do not find either of the hexagonal phases, $MgZn_2$ or $MgNi_2$, to be stable in the range of parameter studied.

In 1999 J. H. Zhu et al. (1999) published a paper putting forward an interpretation for the observed range of size ratios found in Laves phases of between 1.05 and 1.67. By plotting the size

Table 7.5: New structures found in our database that have cubic symmetry and are not present in the ICSD. Row colours (key below) are based on findings from our database where the condition applies if true at any point in parameter space.

Label	Composition		Space group	Example locations
$\eta 1$	A_3B_4		Im3 #217	
$\eta 2$	AB_6		$Im\bar{3}m$ #229	Figure 7.19b
$\eta 3$	AB_8		$Fm\bar{3}m$ #225	Figure 7.19a
$\eta 4$	AB_8		$Fm\bar{3}m$ #225	Figure 7.19a
$\eta 5$	AB		Immm #71	

A B Stable Ground state (off hull) Metastable

ratio against formation enthalpy for 155 different compounds they find that the formation enthalpy becomes positive outside this range. Furthermore the minimum of the formation enthalpy occurs at the ideal size ratio. We have reproduced this from our data on these three structures in fig. 7.21. The minimum falls as expected and as found by J. H. Zhu et al. (1999) there are two different energy gradients to either side. The explanation given for this is that for ratios less than the ideal, B particles (atoms in Zhu's case) are being squashed by As with some associated energy penalty and vice versa with a different penalty for larger ratios. This is an example of a case where our database can be used to quickly test a hypothesis and give insight into more than just ground state by issuing a simple query.

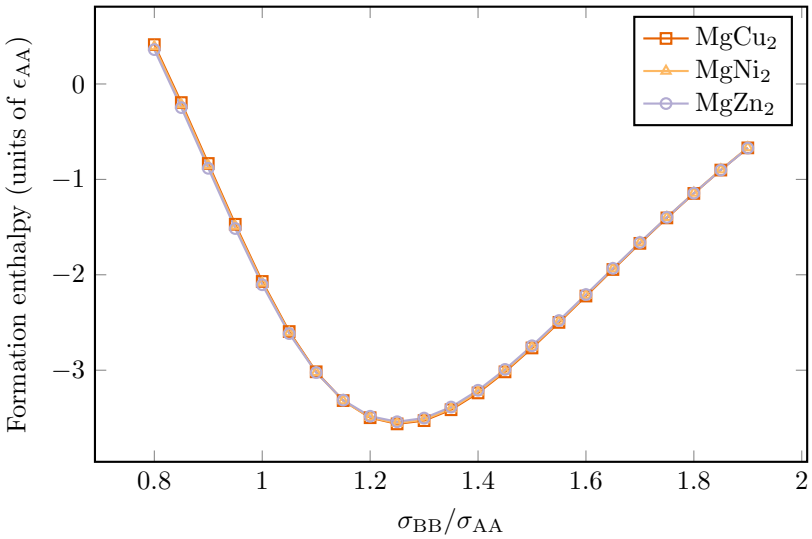


Figure 7.21: A plot of the formation enthalpy for the three Laves phases as a function of the σ_{BB} length scale with σ_{AB} assuming the mean ($\frac{1}{2}(\sigma_{AA} + \sigma_{AB})$) and $\epsilon_{AB} = 1.5, \epsilon_{BB} = 0.5$.

One final observation arising from our plot is that the range of negative formation enthalpy in our results is significantly larger than that found J. H. Zhu et al. This is likely to be a result of our choice of energy scales which determine the slopes either side of the minimum. This raises the possibility that it may be desirable to use a set of ϵ_{ij} values that reproduce the correct range in the hope of more closely capturing energy ordering of close-packed intermetallic structures, in effect fitting the potential to formation enthalpy.

7.5.3 How much of nature is in Lennard-Jones?

To gauge just how broad the range of structures captured by the Lennard-Jones system is we extract the relevant prototypes from the ICSD and check how many of these are also found in our prototypes collection. To form a fair comparison we keep only prototypes that a) have one or two species, b) are ordered, and c) have less than 8 A or B atoms in the primitive unit cell. In total there are 431 such prototypes spanning 20 stoichiometries in the 2012 edition of the ICSD.

Figure 7.22 shows the results broken down by stoichiometry. In total 199 out of 431 prototypes are found in the binary Lennard-Jones system. This results is even more impressive when one

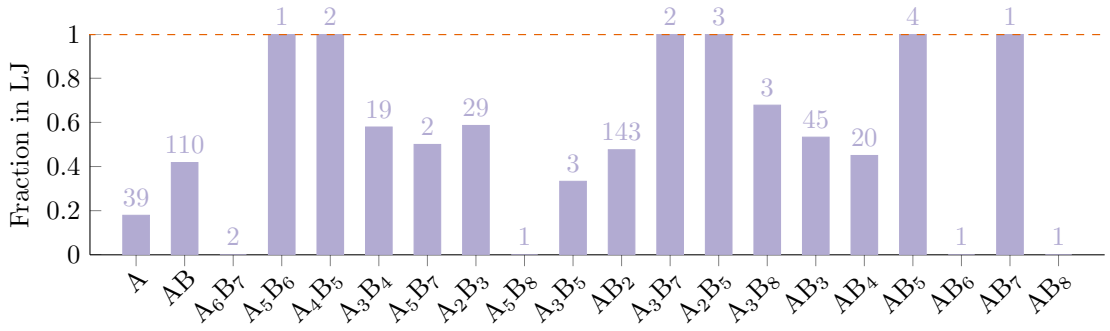


Figure 7.22: A comparison of what fraction of prototypes found in the ICSD are also found somewhere in the binary Lennard-Jones system. Numbers above bars indicate the total number of prototypes of that stoichiometry that were compared against.

considers the number of ICSD entries this represents. From the total of 27,406 structures tagged with one of the 431 prototypes a full 23,722 correspond to the group captured by Lennard-Jones. Thus in total over 86% of the relevant set of structures can be reproduced.

An interesting question for future work is to ask if there are any commonalities amongst the structure types not captured. Is it simply the isotropic character of Lennard-Jones or something else that precludes it from capturing these. A full list of both the found and missing structure is given in appendix A.1.

7.6 Applications

7.6.1 Self-assembly potential design

Self-assembly has been receiving ever increasing attention in large part because of its potential as a micro- or nano-fabrication process capable of producing complex, technologically interesting, structures. A fundamental step to realising this potential is to understand the connection between the interactions of assembling constituents and the assembled structure. Put simply we must have an answer to the question: What effective interaction potential is needed to assemble structure X?

Considerable progress has been made on the so called inverse problem of designing an interaction potential to assemble a particular structure. A notable example is the work of Torquato and co-workers (Rechtsman, Stillinger and Torquato, 2005, 2006a) who use an optimisation scheme to tune a parametrised monatomic isotropic potential for a particular target structure. They were able to virtually assemble a range of structures including cubic (Rechtsman, Stillinger and Torquato, 2006b), wurtzite (Rechtsman, Stillinger and Torquato, 2007) and diamond (Marcotte, Stillinger and Torquato, 2013). Open structures such as these are particularly challenging to reproduce with isotropic potentials which often favour close-packed arrangements. Jacobi and co-workers also tackled open structures in 2D including square and Kagomé (Edlund, Lindgren and Jacobi, 2011a,b) lattices. They take a different approach to Torquato et al. Rather than optimising a parametrised potential they use the reciprocal space representation to directly generate a potential

in a way that guarantees that the lattice is the unique ground state.

While impressive in their ability to reproduce low coordinated structures using isotropic potentials, inverse design methods often generate complicated potentials with several nodes and subtle features that are unlikely to be reproducible in experiment. This shortcoming is of course not lost on the authors who often make explicit attempts to simplify their potentials without affecting the ground state. Our approach is different. Rather than complicating the potential we stick with Lennard-Jones but allow two (or potentially more) particle species. This feels much more in tune with the way that structural diversity arises in nature. For example most of the elements adopt either *bcc*, *fcc* or *hcp* crystal structures with only mixtures accessing a wider range of structure types. Those elements that do crystallise into less common structures are almost exclusively non-metallic exhibiting directional covalent bonds. Here again the occurrence of non-simple structures can be seen as a product of the presence of two species, in this case protons and semi-localised electrons.

From our maps finding the appropriate potential parameters is trivial. One need only find the structure type of interest (or issue a database query) and read off the corresponding values. Furthermore, the following additional information is immediately available which would need to be calculated explicitly when using inverse design methods.

1. The maps give an indication of the range of parameter space range over which each structure is stable. Practically this can be used to estimate the tolerance to which interaction must be realised in experiment to assemble that structure.

2. Because the low lying portion of the energy landscape is converged, metastable structures close in energy to the ground state are also contained in the database. If interactions can be tuned with some fidelity it may be desirable to choose the set of parameters with the largest energy gap to the first metastable structure to maximise the crystallisation rate of the target structure.

3. Stoichiometry maps provide information about the compositional stability of particular phases in addition to phase stability information.

The value of our approach to structure assembly should by now be clear, however Edlund, Lindgren and Jacobi question the practicality of such an approach: “By combining and tuning forces including, but not limited to, van der Waals, electrostatic, magnetic, and steric forces as well as forces of purely entropic origin, it is possible today to construct complex effective interactions. With such a profusion of possibilities, the forward statistical mechanics prescription of sweeping the parameter space to give a complete picture of the possible structures becomes computationally infeasible, and the inverse formulation increasingly appealing.” Our claim is that it is unnecessary to resort to such a sophisticated treatment for many self-assembling systems and in fact that

Lennard-Jones alone is a good approximation to the interparticle interactions. In what follows we provide *a posteriori* evidence to support this claim by showing that experimental structures appear in the portions of parameter space expected based on the properties of the assembling constituents.

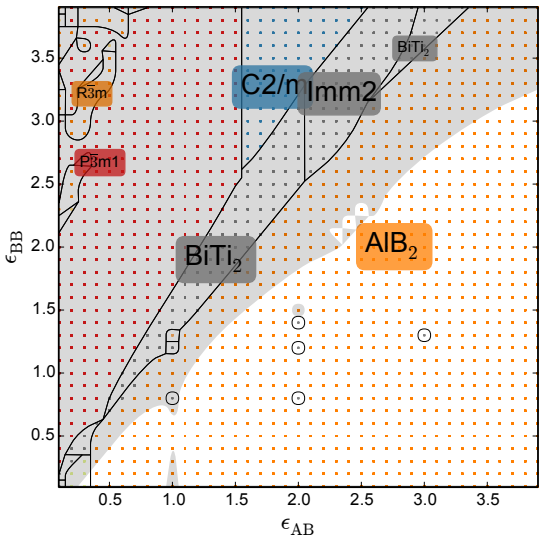
Interaction potential

It can be difficult to characterise the interaction potential of self-assembling constituents however some generalisations can be made based on the available evidence. Experiments involving sub-micrometer colloidal particles are typically performed with a solvent whose refractive index is matched to that of the particles to suppress the van der Waals attraction leaving an effective hard-sphere potential (P. Bartlett, Ottewill and Pusey, 1992; Hunt, Jardine and P. Bartlett, 2000; Pusey et al., 1989). Sterically stabilised nanocrystals tend to be more soft-sphere like with van der Waals attractions and steric repulsion at short range. Both experiment (Shah et al., 2002) and theory (Rabani and Egorov, 2001) suggest that these systems have Lennard-Jones like, single minimum, interaction potentials. As shown in table 7.1 charged or magnetic constituents expand the range of accessible structures, however these systems are less likely to be well approximated by the Lennard-Jones potential alone.

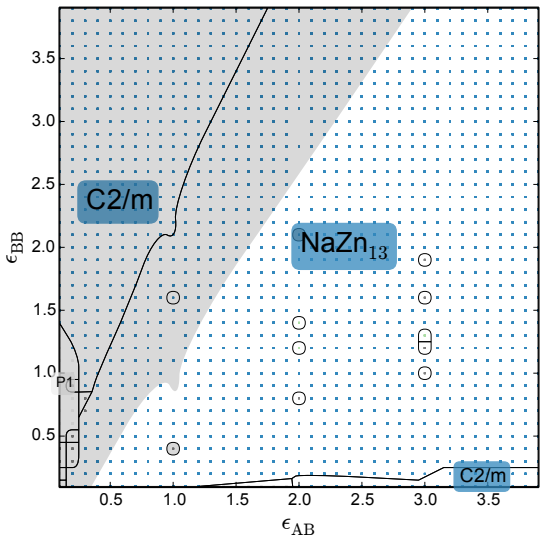
Validation

Much of the current thread of work on self assembled structure of hard and soft spheres begins with an analysis by Sanders (1980) on an usual sample of gem opal found in Brazil. Opal is composed of silica spheres with typical diameters between 0.15 and 0.4 μm. To a first approximation these can be considered to be hard and non-interacting. Ordinary opal has an amorphous structure, however under certain growth conditions the spheres become close packed forming microcrystalline gem opal. The microcrystals diffract light making the gem appear to change colour when changing the angle of illumination. Typically, a sample is composed of a single size of sphere however the sample analysed by Sanders was found to contain two distinct sphere sizes with a size ratio of $r_L/r_S = 0.58$. These crystallised into two stiochiometries, AB_2 and AB_{13}, which were identified to be isostructural with AlB_2 and $NaZn_{13}$ respectively. As can be seen in table 7.1 many of the subsequent, lab-based, self-assembly studies go on to find the same structures.

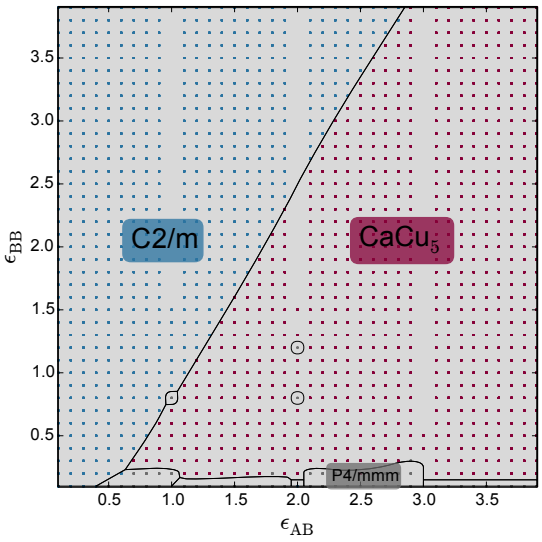
Figure 7.23 shows maps with length scales set to those identified by Sanders and energy scales allowed to vary. For both AB_2 and AB_{13} we find the same structures as Sanders. Furthermore, as expected, they are compositionally stable roughly in the regions were $\epsilon_{AB} > \epsilon_{BB}$. For AB_5 we also find the same structure $CaCu_5$ as found by Shevchenko, Dmitri V. Talapin et al. (2002) for this stoichiometry. These verifications lend support to our assertion that in terms of ground states the Lennard-Jones potential is a good approximation to self assembling systems with either hard sphere or simple soft sphere interactions.



(a) AB_2



(b) AB_{13}



(c) AB_5

Figure 7.23: A portion of parameter space with size ratios set to that of the silica spheres found by Sanders (1980): $\sigma_{AB} = 0.79$, $\sigma_{BB} = 0.58$.

7.6.2 Structure prediction

Structure prediction by ionic substitution has been, at least in terms of sheer numbers, perhaps the most successful prediction technique to date. Long before the advent of accurate modelling of inorganic system chemists used heuristic such as the Goldschmidt rules (Goldschmidt, 1926) to identify ions that are likely to successfully substitute for one and other allowing them to postulate and synthesise new compounds. Such practice remains common today, for example the discovery of superconducting pnictide oxide $LaFeAsO_{1-x}F_x$ prompted chemists to attempt synthesis of analogues with lanthanum replaced by other rare-earth elements (Johrendt and Pöttgen, 2008). The advent of accurate modelling methods, primarily DFT, has only expanded the prevalence of this prediction method, now often being performed in an automated data mining fashion (C. C. Fischer et al., 2006; Hautier, C. C. Fischer et al., 2010). Ionic substitution rules have also been improved

upon using data mining (Hautier, C. Fischer et al., 2011) enabling more accurate initial guesses to be made.

Of course, a necessary prior for all predictions by ionic substitution is a set of known structures. Knowing which ones to choose and which ions to substitute is perhaps destined to become largely automated, however this basic limitation will still exist. This means that while the number of successful predictions made may be high they will not extend to the truly new structures such as the fullerenes, graphene or quasicrystals deserving of Nobel prizes. This drawback does not apply to prediction methods based on global optimisation, however they have their own limitations including the exponential scaling in number of minima of the underlying landscape.

We propose using our database as a structure encyclopedia to be combined with structure prediction by ionic substitution. In a way this acts as a bridge of sorts between the two previously mentioned prediction methods. The presence of a wide range of new structure in our database opens up the possibility of finding something completely new without hitting the scaling wall of global optimisation. Using this hybrid approach predictions such as that of Müller and Zunger (2001) and later Hart (2007) involving truly new structure will surely become all the more common.

7.7 Beyond the periodic table

The periodic table imposes a wealth of constraints on our ability to form structures. It is clear looking back to fig. 7.1 (or from knowledge of high school chemistry) that atoms with larger covalent radii at the bottom left of the table tend to be simultaneously more electronegative and vice versa for small atoms at the top right. In fig. 7.24 we plot the radius and electronegativity ratios for most element pairs. The vast majority are found in the radius ratio range $0.25 \rightarrow 2$ and electronegativity ratio range $0.2 \rightarrow 2.5$. This fact has naturally biased our knowledge of structure to this relatively small window, and that is to ignore constraints on other structure determining properties imposed by nature.

The ability to assemble superstructures from constituents with tunable properties holds promise for breaking free from the periodic table and exploring a wealth of as yet unknown structures with potentially interesting and useful properties. The diversity of available building blocks has grown enormously in recent years, both in terms of sizes and interaction properties. At the bottom of the size range one can consider atoms of nobel gases whose diameters span from 0.064 nm in the case of helium up to 0.260 nm for xenon (ignoring radon). For example the predicted high-pressure structure of XeC_8 by Oganov, Ma et al. (2007) can also be found in the Lennard-Jones database, with xenon creating relatively large channels between the carbon atoms similar in structure to those shown in table 7.8. Moving up in scale it is possible to create so called shape-persistent molecular nanoparticles (MNPs) that mimic individual atoms. Their sizes can vary from between 1 and 30 nm (Tomalia, 2009) with a variety of interaction and surface properties possible. These in

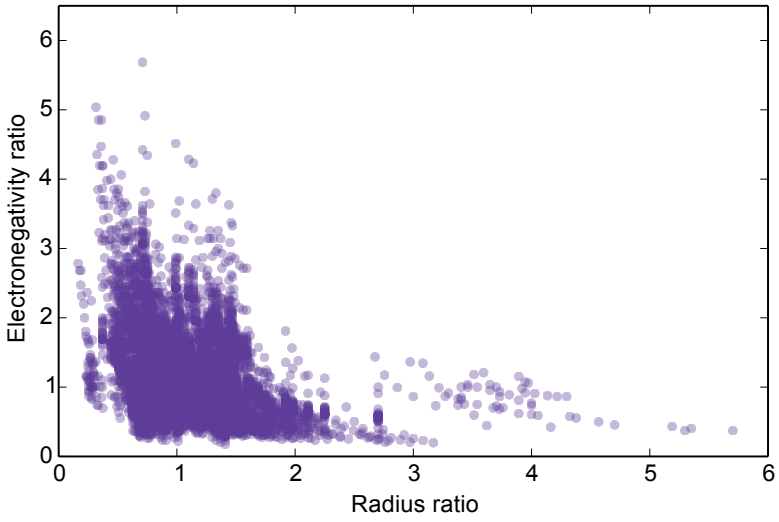


Figure 7.24: A plot of the Pauling electronegativity ratios against covalent radius ratios for most element pairs in the periodic table. The bulk fall in the radius ratio range $0.25 \rightarrow 2$ and electronegativity ratio range $0.2 \rightarrow 2.5$. However, systems not based on atoms as units have no such limitations and can therefore access a wider range of structures.

turn can be used in a hierarchical fashion to create giant polyhedra and supramolecular assemblies in a controlled way, a recent overview of which can be found in W. B. Zhang et al. (2014).

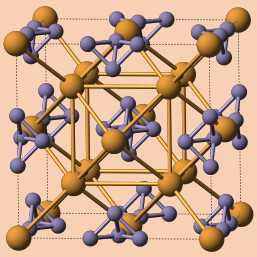
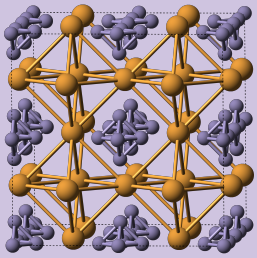
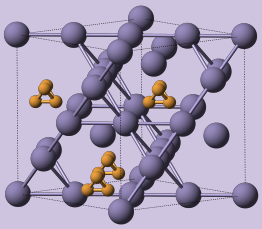
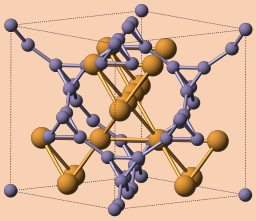
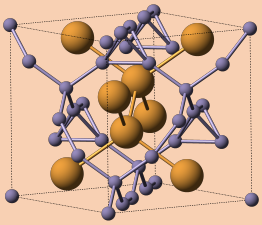
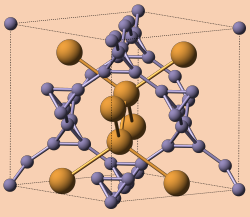
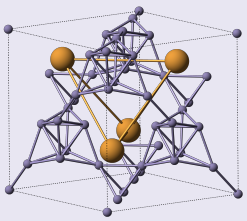
Of particular interest to us are MNPs used to assemble crystal structures. Perhaps the most developed examples of such systems are provided by Roy and co-workers (X. Roy et al., 2013; Turkiewicz et al., 2014). They use a variety of clusters that act as *superatoms* including C_{60} (7.1 Å), $Co_6Se_8(PEt_3)_6$ (8.5 Å), $Cr_6Te_8(PEt_3)_6$ (9.0 Å) and $Ni_6Te_8(PEt_3)_8$ (9.2 Å) where Et are ethane molecules used to terminate the phosphorus atoms. They find that $[Co_6Se_8(PEt_3)_6][C_{60}]_2$ and $[Cr_6Te_8(PEt_3)_6][C_{60}]_2$ assemble into the CdI_2 structure while $[Ni_9Te_6(PEt_3)_8][C_{60}]$ forms rocksalt. The cohesive energy comes from van der Waals interactions and from charge transfer between clusters. These hierarchical crystals were shown to exhibit activated electronic transport and in the case of the rocksalt structure, a magnetically ordered phase at low temperature. Furthermore they show that the electronic properties can be tuned by modification of outer ligands without affecting the overall structure. This important work demonstrates the kinds of new possibilities that are afforded in designing custom superatoms that get around the limits imposed by the periodic table.

In what follows we look at some of the structures in our database from regions in parameter space that are outside the range achievable using elements from the periodic table.

7.7.1 Cubic structures

We begin by looking for prototypes with cubic symmetry (space groups #195-#230) with large radius ratios and no analogue in the ICSD. A selection can be seen in table 7.6.

Table 7.6: Structures with cubic symmetry from beyond the periodic table where the A:B size ratio exceeds that typically found in nature. Row colours (key below) are based on findings from our database where the condition applies if true at any point in parameter space.

Label	Composition		Space group	Example locations
$\beta 1$	AB_2		$Fm\bar{3}m$ #225	Figure 7.19e
$\beta 2$	AB_2		$Pm\bar{3}m$ #221	
$\beta 3$	A_4B_5		$F\bar{4}3m$ #216	Figure 7.20a
$\beta 4$	AB_2		$Fd\bar{3}m$ #229	Figure 7.19e
$\beta 5$	A_2B_5		$F\bar{4}3m$ #216	
$\beta 6$	AB_4		$Fd\bar{3}m$ #227	
$\beta 7$	AB_8		$F\bar{4}3m$ #216	

A B Stable Ground state (off hull) Metastable

β1 The large A particles assume a series of tetrakis hexahedra, formed by placing a square pyramid on each face of a cube. Inside each of these is a tetrahedron of B particles with vertices pointing out towards the tips of each square pyramid.

β2 The large A particles form a series of connected tetrahedra, the voids between which are occupied by tetrahedra of B particles.

β3-7 These structures form a series with a common motif. The B particles form the frame of a diamond structure while the A atoms fill the voids in a variety of ways. The amount of B particles can be increased in certain steps forming a more dense frame. β3 is somewhat exceptional in that the A particles are smaller and pack as tetrahedra in the voids left by B particles. In β4 the A particles form a three dimensional Kagomé net. In β5 and β6 the A particles arrange into a second, interpenetrating, diamond structure, filling the voids. Finally in β7 the A particles form an intersecting 3D triangulated lattice.

7.7.2 $\sigma_{\mathbf{AB}} > \sigma_{\mathbf{AA}}, \sigma_{\mathbf{BB}}$

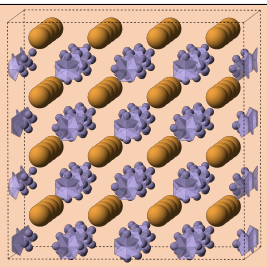
One portion of parameter space the periodic table does not readily extend to is the case where unlike (A-B) length scales are larger than either of the other two (A-A and B-B). The opposite is quite possible for example in cases where the A-B attraction is stronger than either A-A or B-B and the particles are deformable then the mixed dimer may have an shorter equilibrium separation than either pure dimer. Conceptually a larger A-B seems rather odd as it would appear that particles A and B 'see' each other as being larger than either of them see themselves. Theoretically and perhaps practically, however, there is no problem with the situation. Macfarlane et al. have demonstrated that it is possible to synthesise particles composed of gold nanocluster cores coated with DNA linkers. Crucially it is the hydrodynamic radii, dictated by the linkers, and not the core radii that determine the stable structure. Furthermore the linkers can be programmed to create self- or non-self- complementary interactions. By using self-complementary linkers it may be possible to access the regime under discussion.

γ1 This structure is a series of icosahedra surrounding a central particle formed by small B particles that fill the voids of a cubic lattice formed by A particles.

τ-family

Table 7.8 shows a family of structures that form in this portion of parameter space. Their structure consists of a series of channels made of B particles, shown into the page, with A particles forming wires that thread through their centres. As the concentration of A atoms is increased the channels distort away from being circular to accommodate their presence. The addition of B atoms causes

Table 7.7: New structures found in the database that cannot be found in the ICSD. Row colours (key below) are based on findings from our database where the condition applies if true at any point in parameter space.

Label	Composition		Space group
$\gamma 1$	AB_{13}		$Fm\bar{3}c$ #226

A B Stable Ground state (off hull) Metastable

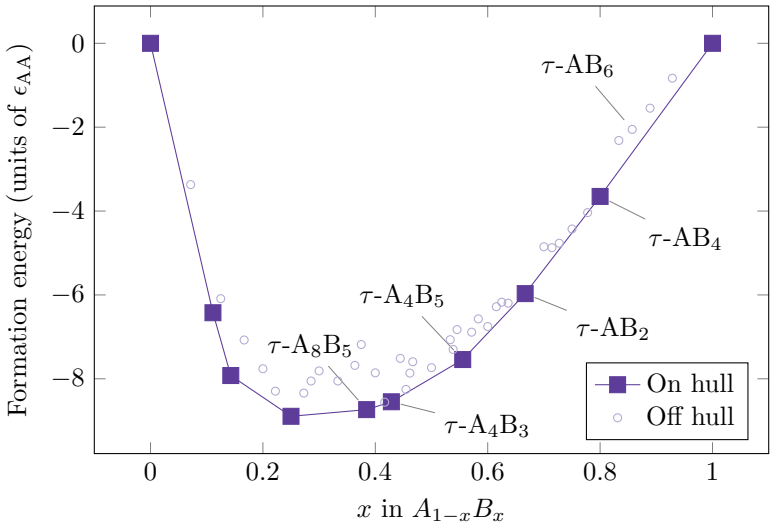
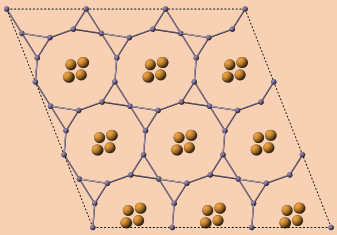
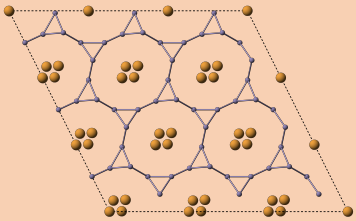
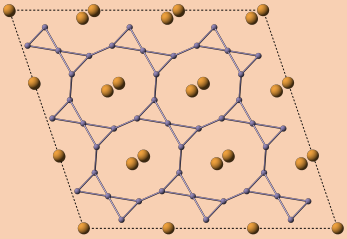
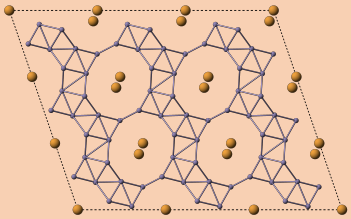
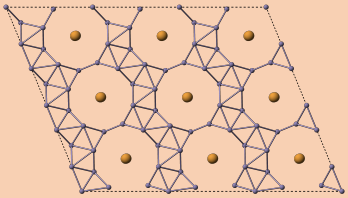
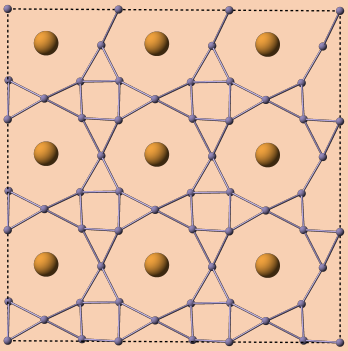


Figure 7.25: A convex hull taken from $\epsilon_{AB} = 1.5$, $\sigma_{AB} = 2.5$, $\epsilon_{BB} = 0.5$, $\sigma_{BB} = 2$. The members of the τ family of structures have been labelled.

Table 7.8: A family of structures from the $\sigma_{AB} > \sigma_{AA}, \sigma_{BB}$ portion of parameter space. The smaller B atoms form a series of channels into the page with A atoms forming wires that thread through their centres. Row colours (key below) are based on findings from our database where the condition applies if true at any point in parameter space.

Label		Space group
$\tau - A_8B_5$		C2/m #12
$\tau - A_4B_3$		Cm #8
$\tau - A_4B_5$		$P\bar{1}$ #2
$\tau - AB_2$		P1 #1
$\tau - AB_4$		Imma #74
$\tau - AB_6$		Imm2 #44

A B Stable Ground state (off hull) Metastable

the walls of the channels to increase in thickness. It is possible to tune the channel diameter by varying the σ_{AB} parameter. Not all members of the family will necessarily be compositionally stable at one parameter point as shown in fig. 7.25. However it is often possible to find other regions of stability for any member of the family.

7.8 Conclusion

Figure 7.4 highlighted just how little of the binary Lennard-Jones system had been studied. By building a high throughput framework combined with a schemaless database we were able to vastly expand on this range both in terms of parameter space and stoichiometry. The result is a collection of structure data and relationships that is much more than simply the product of a giant stamp collecting exercise proving to be of both fundamental and practical value.

On the fundamental side our work continues in the spirit of Hume-Rothery, Laves, Frank and Kasper, Villars, Pettifor and many others who studied the factors affecting structure stability. Rather than trying to correlate factors to structure we have come at the question from the other side and have shown what structures can be formed *given* a small set of factors to vary. The range of structures types found has been huge. We have shown that over 86% of the relevant structures in the ICSD are contained in our database. Perhaps more interesting are those structures that have no known analogue in nature. There are many such examples in our data and as pointed out by Hart having a procedure for identifying those that are more likely to have physical manifestations is key when dealing with so many structures. This may be a fruitful avenue for future work. In addition to the structures themselves our results can provide qualitative insights such as the example given in section 7.5.2 which mirrored results found by J. H. Zhu et al. in explaining the relationship between radius ratio and stability of the Laves phases.

Of the practical applications discussed perhaps most exciting is the connection to the growing field of self-assembly. We have demonstrated that the binary Lennard-Jones system is a good approximation for a range of self-assembling systems reproducing the found ground states in a handful of cases. By breaking free from the constraints imposed by the periodic table it is possible to assemble a much wider range of structures than typically seen in nature bringing with them the possibility for yet unknown and interesting properties. We also speculate that it may be possible to form certain large radius ratio structures in atomic systems by, for example, using C_{60} units in place of a single particle type in our systems. An example of this already exists in the case of the Cr_3C_{60} structure assembled by Macfarlane et al. Needless to say there are obstacles to realising many of our structures. For example, we completely ignore kinetic considerations which are likely to play a large role in determining which phases are accessible. Despite this advances in experimental techniques are maturing and overcoming problems such as jamming making the process more reliable. Theory is also helping to suggest more robust protocols such as the work

of Jacobs, Reinhardt and Frenkel (2015) who discuss methods for achieving optimal nucleation kinetics.

This chapter began with a quote from Feynman suggesting that the most concise piece of information that could be communicated to a primitive society about nature is that all things are made of atoms, attracting each other at long range and repelling at short. Our work should leave no doubt that not only is there a great deal of truth to this claim, but that we ourselves remain ignorant of many of its natural consequences.

8 | Conclusion

. . . at this moment, here we face a critical branch point in history, what we do with our world, right now, will propagate down through the centuries and powerfully affect the destiny of our descendants, it is well within our power to destroy our civilization and perhaps our species as well. . . But we are also capable of using our compassion and our intelligence, our technology and our wealth to make an abundant and meaningful life for every inhabitant of this planet. To enhance enormously our understanding of the Universe, and to carry us to the stars.

Carl Sagan, 1980

2014 was the international year of crystallography with regional projects spanning 80 countries, from Algeria to Vietnam, in celebration of the centenary of the discovery of x-ray diffraction in crystals. It would be difficult to overstate just how much our understanding of the structure of materials has progressed in this relatively short time. In this thesis we have brought together three important tools that have made invaluable contributions on the theory side. *Ab initio* structure prediction is now a mature workhorse able to tackle much more than simple, small unit cell, materials. We have tackled three such examples including high-pressure multicomponent earth core materials, complex interactions in the phosphorus and arsenic systems and large unit cell grain boundaries in graphene. On the other hand rationalisation got us a long way in understanding certain classes of system, for example intermetallics and normal valence compounds. In this work we have used big data as a bridge between the two to gain a more complete picture. Along the way we have generated a wealth of possible structures, including many not considered before because they lie outside the reach of the periodic table.

The understanding of structure is interesting in the pure sense but for the field to be useful it is important to ask questions about structure-property relationships. Certainly big data and sophisticated tools such as machine learning will have an ever expanding role to play here and this would be an interesting avenue for future work. If, for example, machine learning can identify connections between structure types and properties our encyclopedia of structures could be used to

select candidates for a particular target property. These could then be screened in high-throughput electronic structure calculations using a range of substituted ion species.

As we enter the next 100 years following von Laue's discovery, the goal of bringing together all the pieces for a general solution to the inverse problem of going from property to structure to synthesis look almost within reach. If realised, the impact on ordinary people's lives could be as significant as the discovery of x-ray diffraction itself and would truly cement this as the age of designed materials.

A | Appendix

A.1 ICSD/Lennard-Jones comparison

Here we list the structures types comparing the ICSD with our binary Lennard-Jones prototypes collection. ICSD designations have been used throughout.

A.1.1 Found prototypes

OsO4	Al2O3	Pa	CS2
La2Ni3	TiO2(tI12)	BiF5	TiI4
SrS2	PbCl2	Nb5N4	MoS2(3R)
TiSi2	Cr3S4	UTe2	CdI2(hP6)
BaSi2	FeS2(cP12)	Na3As	HgO2
GeS	FeS2(aP12)	Ag2F	CaF2
Rh3Te8	Bi2Te3	PuF3	Litharge
Dy3Ni2	TiO2(tP6)	MoS2(4H)	PbO(oS8)
UCl3	Pd4Pu3	PbO2-alpha	NaCl
Li5Tl2	ZrCl	VO2	Au3K2
Ti5Te4	ZrAs2	Pd2As	Ta2Se
CaSi2(hR3)	BiI3(P3-1m)	Fe2B	Sr2N
Cu3P	C5Nb6	Mo2N	Al4C3
Mo3Se4	HgI2(oS12)	NdBr3	ZrO2(HT)
NbAs	WC	UO3	Si2U3
FeF3	KrF2	Ga2O3	RhSi
Cr2S3(hR10)	Pd4Se	FeSi	Fe2P/K2UF6
TiAs	ZrCl3	MoP2	Fe3N1+x
SnF4	MoB	KO2	UGe2
In2Se3	Mo2B5	U4S3	Rb2Te2
GeAs	ReB3	Cr3B4	LaI

CdCl2(3R)	Pd3Te2	Hg2Pt	Al3Ti
Li2Ga	FeSe(tP2)	IrV	AuBe5
Fe2Si(HT)	CdI2	LiSn	ThSi2
V7O3	TeO2	MgZn2	PdF2(cP12)
LaSi	GeTe(supercell)	MoNi4	Cu2Mg
CdI2(hR6)	AlCe	MoPt2	CuZr2
FeS2(oP6)	Al2Li3	Pd5Ti3	HgMn
Nb2Se3	Al8Mo3	AlLi	Ni3Sn
GeTe(subcell)	Al3Os2	MoO2	Al3Zr4(P6/mmm)
Sn(tI2)	ZrSi2	PtTe	InMg2
MnP	AlTh	SiC(4H)	InNi2
Ca7Ge	Al3Zr2	Cr3C2	Cu3Ti
Y5S7	Ag2Nd	LiPt	Al3Ni2
Re3B	BaCu	Au7Rb3	NaTl
CrB	BiIn	AlDy	Hg4Pt
Nb4N3+x	CeCu2	EuGa	Al3Zr4
HfO2	BCl3	EuIn4	NiHg4
Po(alpha)	TiS	Al3Zr	NdCo2
Fe2C	TlI	Cr3Si	LaF3
CrN	Ag3Sb	CO2(cP12)	Co2Si
Nb2O5(HP)	ZnS(2H)	BiF3	FeB
CaC2	Sm	Th3P4	CaCl2
NiO	BiI3	MoSi2	CeGe5
AgCl(mP4)	La	NiAs	EuGe2/PtTe2
La2O3	CuMg2	Al2Cu	Al4U
Cu2O	CuTi	AuCu3	IrN2
TaS2(6R)	Cu4Ti3	Cu	CaCu5
CaSi2(hR6)	ZrGa2	W	
ZnS(cF8)	HgK	CsCl	

A.1.2 Missing prototypes

VO2(HT)	BiI(mS16)	CrP2	Au2O3
ZnAs	P4S4	CrP4	GaTe
CaSb2	U3Si	SrSi2	Si3N4
PdF4	Fe2Tb	CuF2	CdI2(hP12)
BiI(mS32)	ZnCl2	AgF2	MoBr3

Gd2Cl3	I2	HgCl	Ga(Cmcm)
NaP	CdI2(hP9,P3m1)	PdS	GaAs
SiF4	Ni4B3	NbTe4	NbS2(oS6)
SnF2	FeSi2	PtI2	KO2(mS12)
NbO	Zn3P2(cP10)	UCl4	Ag2S-LT
MoTe2	CdP4	SF6	Pu(oF8)
V4As3	Be3N2	AuTe2	Te(mP4)
InS	TaN	NiS	Graphite(2H)
CrO3	Nb2Se	AlCu	WO3(TP8)
SnP3	CdP2	TaS2(hP6)	Quartz high
Nb3Te4	TaSe2(P63mc)	TlTe	Si(oS16)
PdS2	LiAs	YS	HgO
U3O8	PdO	LaS	Sb(mP4)
CrSi2	XeF2	ReSe2	Na2C2
SiS2	XeF4	FeS(2H)	Fe8N
Si(HP)	H2S	CaP3	C(cF8)
TaSe2	S6	Si3N4(P63/m)	ZrSb
BeO(tP8)	RhBr3	SiO2(P1)	Cu2Te(HT)
MnF3	NaN3(alpha)	CuCl	Cristobalite high
Cristobalite(alpha)	BiSe	LiSi	Ti2C
CoAs	Se(gamma)	Er3Ge4	Pt2Si
SiC(6H)	Quartz low	KSb2	FeGe
Massicot	Ni3S2	SiO2(mS24)	Mg
SiO2(C2221)	CdI2(hP9)	HgS	Graphite(3R)
CuO(mS8)	Ti8C5	P2I4	Sn
GaS	SiS2(Ibam)	Tridymite	As
SiTi	Cu Br	Ta2H(oS6)	TbGe2
HgI2	PdCl2	NS	ZrH
HgCl2	BrF3	BaS2	Rb2C2
BaP3	RuB2	CdI2(hR9)	CaC2(mS12)
BaS3	Li3N	CdI2(hR12)	Bi(I4/mcm)
Ru2B3	CuP2	CuTe	Bi(III)
WB2	BN	SmSb2	AgSr
HgF	ScF3	Po(hR1)	CaIn2
P(black)	Se2Cl2	NaSi	CaHg2
ReB2	K4P6	Co5Ge7	Cd2Ce
NdTe3	SiO2(aP9)	CoSb2	V2O3

U	Pd5Th3	MnO2	CaO(hP4)
ZrO2(mP12)	Pd3Ti2	CO	BN(mS16)
NbTe4(P4cc)	Re2U	La2Sb	Sc10+x
ThC2	SrSn4	Pt3O4	(Ca8)xCa2
AlAu	Bi	CoO	FeSe
CoU	Tb6Br7	In	Ge(hP8)
Fe7W6	NbTe2	Ga	Ge
Ga3Pt5	Ag2Te	ScSn2	LiSn(tI24)
HfGa2	CuCl2	CeTe3	ZnS(15R)
Hg(LT)	Ga2S3	Sr(guest atoms)	ZnS(21R)
HgIn	Pt3Ge	GdSn3	ZnS(hR16)
HgNa	Fe3Pt	PtSn4	Mn3As
IrTa	IrU	Co2Nd	BaPt
Ni3Sn4	Cs(tP8)	InAs	CoSn
NiTi	Ag2O3	K20-x	Ag2Se
Pb4Pt	ReO3	RhN2	

Bibliography

Abbas, Sayeed and Lodge, Timothy P (2008). 'Superlattice formation in binary mixtures of block copolymer micelles.' EN. In: *Langmuir : the ACS journal of surfaces and colloids* 24.12, pp. 6247–53. ISSN: 0743-7463. DOI: 10.1021/la8002367.

Ackerbauer, Sarah et al. (2009). 'The constitution of the ternary system Fe–Ni–Si'. In: *Intermetallics* 17.6, pp. 414–420. ISSN: 09669795. DOI: 10.1016/j.intermet.2008.11.016.

Akahama, Yuichi, Kobayashi, Mototada and Kawamura, Haruki (1999). 'Simple-cubic–simple-hexagonal transition in phosphorus under pressure'. In: *Physical Review B* 59.13, pp. 8520–8525. ISSN: 0163-1829. DOI: 10.1103/PhysRevB.59.8520.

Alfè, D, Gillan, M. J. and Price, Geoffrey D. (2002). 'Composition and temperature of the Earth's core constrained by combining ab initio calculations and seismic data'. In: *Earth and Planetary Science Letters* 195.1-2, pp. 91–98. ISSN: 0012821X. DOI: 10.1016/S0012-821X(01)00568-4.

Allen, Frank H (2002). 'The Cambridge Structural Database: a quarter of a million crystal structures and rising.' In: *Acta crystallographica. Section B, Structural science* 58.Pt 3 Pt 1, pp. 380–8. ISSN: 0108-7681.

Allmann, Rudolf and Hinek, Roland (2007). 'The introduction of structure types into the Inorganic Crystal Structure Database ICSD.' In: *Acta crystallographica. Section A, Foundations of crystallography* 63.Pt 5, pp. 412–7. ISSN: 0108-7673. DOI: 10.1107/S0108767307038081.

An, Jinho et al. (2011). 'Domain (grain) boundaries and evidence of "twinlike" structures in chemically vapor deposited grown graphene.' In: *ACS nano* 5.4, pp. 2433–9. ISSN: 1936-086X. DOI: 10.1021/nn103102a.

Anderson, Don L. (1989). *Theory of the Earth*. Boston, MA: Blackwell Scientific Publications, pp. 289–305. ISBN: 0865423350.

Antonangeli, Daniele et al. (2010). 'Composition of the Earth's inner core from high-pressure sound velocity measurements in Fe–Ni–Si alloys'. In: *Earth and Planetary Science Letters* 295.1-2, pp. 292–296. ISSN: 0012821X. DOI: 10.1016/j.epsl.2010.04.018.

Aurenhammer, Franz (1991). *Voronoi diagrams—a survey of a fundamental geometric data structure*. DOI: 10.1145/116873.116880.

Bachhuber, Frederik et al. (2014). 'The extended stability range of phosphorus allotropes.' In: *Angewandte Chemie (International ed. in English)* 53.43, pp. 11629–33. ISSN: 1521-3773. DOI: 10.1002/anie.201404147.

Badro, James, Côté, Alexander S. and Brodholt, John P. (2014). 'A seismologically consistent compositional model of Earth's core.' In: *Proceedings of the National Academy of Sciences of the United States of America* 111.21, pp. 7542–5. ISSN: 1091-6490. DOI: 10.1073/pnas.1316708111.

Barlow, W. M. (1883). *Probable Nature of the Internal Symmetry of Crystals*. DOI: 10.1038/029186a0.

Barrett, C. S. and Massalski, T. B. (1980). *Structure of Metals: Crystallographic Methods, Principles and Data*. 3rd. Pergamon. ISBN: 0080261728.

Bartlett, Paul (1990). 'A model for the freezing of binary colloidal hard spheres'. en. In: *Journal of Physics: Condensed Matter* 2.22, pp. 4979–4989. ISSN: 0953-8984. DOI: 10.1088/0953-8984/2/22/018.

Bartlett, Paul, Ottewill, R. H. and Pusey, P. N. (1990). 'Freezing of binary mixtures of colloidal hard spheres'. In: *The Journal of Chemical Physics* 93.2, p. 1299. ISSN: 00219606. DOI: 10.1063/1.459142.

— (1992). 'Superlattice formation in binary mixtures of hard-sphere colloids'. In: *Physical Review Letters* 68.25, pp. 3801–3804. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.68.3801.

Barzilai, Jonathan and Borwein, Jonathan M. (1988). 'Two-Point Step Size Gradient Methods'. In: *IMA Journal of Numerical Analysis* 8.1, pp. 141–148. ISSN: 0272-4979. DOI: 10.1093/imanum/8.1.141.

Baudler, Marianne and Glinka, Klaus (1993). 'Contributions to the chemistry of phosphorus. 218. Monocyclic and polycyclic phosphines'. In: *Chemical Reviews* 93.4, pp. 1623–1667. ISSN: 0009-2665. DOI: 10.1021/cr00020a010.

Baus, M (1987). 'Statistical mechanical theories of freezing: An overview'. In: *Journal of Statistical Physics* 48.5-6, pp. 1129–1146. ISSN: 0022-4715. DOI: 10.1007/BF01009537.

Beghein, Caroline and Trampert, Jeannot (2003). 'Robust normal mode constraints on inner-core anisotropy from model space search.' In: *Science (New York, N.Y.)* 299.5606, pp. 552–5. ISSN: 1095-9203. DOI: 10.1126/science.1078159.

Belsky, Alec et al. (2002). 'New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design'. In: *Acta Crystallographica Section B Structural Science* 58.3, pp. 364–369. ISSN: 01087681. DOI: 10.1107/S0108768102006948.

Berger, Lev I. (1996). *Semiconductor Materials*. CRC Press, pp. 81–86.

Berthelot, Daniel (1898). 'Sur le mélange des gaz'. In: *Comptes rendus hebdomadaires des séances de l'Académie des sciences* 126, pp. 1703–1855.

Blerdsoe, W. W. (1961). 'The use of biological concepts in the analytical study of systems.' In: *ORSA-TIMS National meeting*. San Francisco, CA.

Blöchl, Peter E. (1990). 'Generalized separable potentials for electronic-structure calculations'. In: *Physical Review B* 41.8, pp. 5414–5416. ISSN: 01631829. DOI: 10.1103/PhysRevB.41.5414.

Böcker, Stephan and Häser, Marco (1995). 'Covalent Structures of Phosphorus: A Comprehensive Theoretical Study'. In: *Zeitschrift für anorganische und allgemeine Chemie* 621.2, pp. 258–286. ISSN: 0044-2313. DOI: 10.1002/zaac.19956210215.

Box, George E P (1957). 'Evolutionary Operation: A Method for Increasing Industrial Productivity'. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 6.2, pages. ISSN: 00359254.

Brazdova, Veronika and Bowler, David R. (2013). *Atomistic Computer Simulations: A Practical Guide*. Wiley. ISBN: 978-3-527-41069-9.

Bremermann, H. J. (1962). 'Optimization through evolution and recombination'. In: *Self-Organizing Systems*. Washington, D. C.: Spartan Books, pp. 93–106. ISBN: 978-1258715038.

Bulgakov, A. V., Bobrenok, O. F. and Kosyakov, V. I. (2000). 'Laser ablation synthesis of phosphorus clusters'. In: *Chemical Physics Letters* 320.1-2, pp. 19–25. ISSN: 00092614. DOI: 10.1016/S0009-2614(00)00234-7.

Bulgakov, A. V., Bobrenok, O. F., Kosyakov, V. I. et al. (2002). 'Phosphorus clusters: Synthesis in the gas-phase and possible cagelike and chain structures'. In: *Physics of the Solid State* 44.4, pp. 617–622. ISSN: 1063-7834. DOI: 10.1134/1.1470540.

Bulgakov, A. V., Bobrenok, O. F., Ozerov, I. et al. (2004). 'Phosphorus cluster production by laser ablation'. In: *Applied Physics A* 79.4-6. ISSN: 0947-8396. DOI: 10.1007/s00339-004-2782-z.

Butler, Sheneve Z et al. (2013). 'Progress, challenges, and opportunities in two-dimensional materials beyond graphene.' In: *ACS nano* 7.4, pp. 2898–926. ISSN: 1936-086X. DOI: 10.1021/nn400280c.

Cartwright, H.M. et al. (2004). *Applications of Evolutionary Computation in Chemistry*. Ed. by Roy L. Johnston. Springer. ISBN: 978-3540402589. DOI: 10.1007/b10607.

Černý, V. (1985). 'Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm'. In: *Journal of Optimization Theory and Applications* 45.1, pp. 41–51. ISSN: 0022-3239. DOI: 10.1007/BF00940812.

Chen, Zhuoying, Moore, Jenny et al. (2007). 'Binary nanoparticle superlattices in the semiconductor-semiconductor system: CdTe and CdSe.' en. In: *Journal of the American Chemical Society* 129.50, pp. 15702–9. ISSN: 1520-5126. DOI: 10.1021/ja076698z.

Chen, Zhuoying and O'Brien, Stephen (2008). 'Structure direction of II-VI semiconductor quantum dot binary nanoparticle superlattices by tuning radius ratio.' EN. In: *ACS nano* 2.6, pp. 1219–29. ISSN: 1936-086X. DOI: 10.1021/nn800129s.

Chung, S. J., Hahn, Th. and Klee, W. E. (1984). 'Nomenclature and generation of three-periodic nets: the vector method'. en. In: *Acta Crystallographica Section A Foundations of Crystallography* 40.1, pp. 42–50. ISSN: 0108-7673. DOI: 10.1107/S0108767384000088.

Clark, Stewart J. et al. (2005). 'First principles methods using CASTEP'. In: *Zeitschrift für Kristallographie* 220.5/6/2005, pp. 567–570. ISSN: 0044-2968. DOI: 10.1524/zkri.220.5.567.65075.

Corbridge, D. E. C. (1974). *Structural Chemistry of Phosphorus*. Elsevier Science Ltd. ISBN: 978-0444410733.

Côté, Alexander S., Vočadlo, Lidunka and Brodholt, John P. (2012). 'Ab initio simulations of iron–nickel alloys at Earth's core conditions'. In: *Earth and Planetary Science Letters* 345-348, pp. 126–130. ISSN: 0012821X. DOI: 10.1016/j.epsl.2012.06.025.

Cottin, X. and Monson, P. A. (1993). 'A cell theory for solid solutions: Application to hard sphere mixtures'. In: *The Journal of Chemical Physics* 99.11, p. 8914. ISSN: 00219606. DOI: 10.1063/1.465560.

— (1995). 'Substitutionally ordered solid solutions of hard spheres'. In: *The Journal of Chemical Physics* 102.8, pp. 3354–3360. DOI: 10.1063/1.469209.

— (1996). 'Solid–fluid phase equilibrium for single component and binary Lennard-Jones systems: A cell theory approach'. In: *The Journal of Chemical Physics* 105.22, pp. 10022–10029. DOI: 10.1063/1.472832.

Cummings, Aron W et al. (2014). 'Charge transport in polycrystalline graphene: challenges and opportunities.' In: *Advanced materials (Deerfield Beach, Fla.)* 26.30, pp. 5079–94. ISSN: 1521-4095. DOI: 10.1002/adma.201401389.

Curtarolo, Stefano et al. (2012). 'AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations'. In: *Computational Materials Science* 58, pp. 227–235. ISSN: 09270256. DOI: 10.1016/j.commatsci.2012.02.002.

Damasceno, Pablo F., Engel, Michael and Glotzer, Sharon C. (2012). 'Predictive self-assembly of polyhedra into complex structures.' In: *Science (New York, N.Y.)* 337.6093, pp. 453–7. ISSN: 1095-9203. DOI: 10.1126/science.1220869.

Dehlinger, Ulrich (1955). *Theoretische Metallkunde*. Vol. 13. Reine und angewandte Metallkunde in Einzeldarstellungen. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-53356-3. DOI: 10.1007/978-3-642-53397-6.

Delhommelle, Jérôme and Millié, Philippe (2001). *Inadequacy of the Lorentz-Berthelot combining rules for accurate predictions of equilibrium properties by molecular simulation*. DOI: 10.1080/00268970010020041.

Denton, A. and Ashcroft, N. (1990). 'Weighted-density-functional theory of nonuniform fluid mixtures: Application to freezing of binary hard-sphere mixtures'. In: *Physical Review A* 42.12, pp. 7312–7329. ISSN: 1050-2947. DOI: 10.1103/PhysRevA.42.7312.

Ding, Yi et al. (2014). 'Anisotropic elastic behaviour and one-dimensional metal in phosphorene'. In: *physica status solidi (RRL) - Rapid Research Letters* 8.11, pp. 939–942. ISSN: 18626254. DOI: 10.1002/pssr.201409385.

Dion, M. et al. (2004). 'Van der Waals Density Functional for General Geometries'. In: *Physical Review Letters* 92.24, p. 246401. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.92.246401.

Downs, Robert T. and Hall-Wallace, Michelle (2003). 'The American Mineralogist crystal structure database'. In: 88, pp. 247–250.

Doye, Jonathan P. K. and Massen, Claire P (2005). 'Characterizing the network topology of the energy landscapes of atomic clusters.' In: *The Journal of chemical physics* 122.8, p. 84105. ISSN: 0021-9606. DOI: 10.1063/1.1850468.

Doye, Jonathan P. K., Wales, David J. and Miller, Mark A. (1998). 'Thermodynamics and the global optimization of Lennard-Jones clusters'. In: *The Journal of Chemical Physics* 109.19, p. 8143. ISSN: 00219606. DOI: 10.1063/1.477477.

Dubrovinsky, L et al. (2007). 'Body-centered cubic iron-nickel alloy in Earth's core.' In: *Science (New York, N.Y.)* 316.5833, pp. 1880–3. ISSN: 1095-9203. DOI: 10.1126/science.1142105.

Duong, Dinh Loc et al. (2012). 'Probing graphene grain boundaries with optical microscopy.' In: *Nature* 490.7419, pp. 235–9. ISSN: 1476-4687. DOI: 10.1038/nature11562.

Edlund, E., Lindgren, O. and Jacobi, M. Nilsson (2011a). 'Designing Isotropic Interactions for Self-Assembly of Complex Lattices'. In: *Physical Review Letters* 107.8, p. 085503. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.107.085503.

— (2011b). 'Novel Self-Assembled Morphologies from Isotropic Interactions'. In: *Physical Review Letters* 107.8, p. 085501. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.107.085501.

Eldridge, M. D., Madden, P. A. and Frenkel, Daan (1993). 'Entropy-driven formation of a superlattice in a hard-sphere binary mixture'. en. In: *Nature* 365.6441, pp. 35–37. ISSN: 0028-0836. DOI: 10.1038/365035a0.

Eldridge, M. D., Madden, P. A., Pusey, P. N. et al. (1995). 'Binary hard-sphere mixtures: a comparison between computer simulation and experiment'. en. In: *Molecular Physics* 84.2, pp. 395–420. ISSN: 0026-8976. DOI: 10.1080/00268979500100271.

Engel, Michael and Trebin, Hans-Rainer (2007). 'Self-Assembly of Monatomic Complex Crystals and Quasicrystals with a Double-Well Interaction Potential'. In: *Phys. Rev. Lett.* 98.22, p. 225505. DOI: 10.1103/PhysRevLett.98.225505.

Fei, Z et al. (2013). 'Electronic and plasmonic phenomena at graphene grain boundaries.' In: *Nature Nanotechnology* 8.11, pp. 821–5. ISSN: 1748-3395. DOI: 10.1038/nnano.2013.197.

Fernández, Julián R and Harrowell, Peter (2003). 'Crystal phases of a glass-forming Lennard-Jones mixture'. In: *Phys. Rev. E* 67.1, p. 11403. DOI: 10.1103/PhysRevE.67.011403.

— (2004). 'Ordered binary crystal phases of Lennard-Jones mixtures.' In: *The Journal of chemical physics* 120.19, pp. 9222–32. ISSN: 0021-9606. DOI: 10.1063/1.1689642.

Filion, Laura and Dijkstra, Marjolein (2009). 'Prediction of binary hard-sphere crystal structures'. In: *Physical Review E* 79.4, p. 046714. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.79.046714.

Fischer, Christopher C. et al. (2006). 'Predicting crystal structure by merging data mining with quantum mechanics.' en. In: *Nature materials* 5.8, pp. 641–6. ISSN: 1476-1122. DOI: 10.1038/nmat1691.

Fletcher, R and Powell, M J D (1963). 'A Rapidly Convergent Descent Method for Minimization'. In: *The Computer Journal* 6.1, pp. 163–168. ISSN: 0010-4620. DOI: 10.1093/comjnl/6.2.163.

Fowler, P. W. and Manolopoulos, David E. (1995). *An Atlas of Fullerenes*. Clarendon Press. ISBN: 0198557876.

Frank, F. C. and Kasper, J. S. (1958). 'Complex alloy structures regarded as sphere packings. I. Definitions and basic principles'. en. In: *Acta Crystallographica* 11.3, pp. 184–190. ISSN: 0365110X. DOI: 10.1107/S0365110X58000487.

— (1959). 'Complex alloy structures regarded as sphere packings. II. Analysis and classification of representative structures'. en. In: *Acta Crystallographica* 12.7, pp. 483–499. ISSN: 0365110X. DOI: 10.1107/S0365110X59001499.

Friedman, G. J. (1956). 'Selective Feedback Computers for Engineering Synthesis and Nervous System Analogy'. Master's Thesis. UCLA.

Friedrich, W., Knipping, P. and Laue, M. (1913). 'Interferenzerscheinungen bei Röntgenstrahlen'. In: *Annalen der Physik* 346.10, pp. 971–988. ISSN: 00033804. DOI: 10.1002/andp.19133461004.

Glass, Colin W., Oganov, Artem R. and Hansen, Nikolaus (2006). 'USPEX—Evolutionary crystal structure prediction'. In: *Computer Physics Communications* 175.11-12, pp. 713–720. ISSN: 00104655. DOI: 10.1016/j.cpc.2006.07.020.

Goldschmidt, V. M. (1926). 'Die Gesetze der Krystallochemie'. In: *Die Naturwissenschaften* 14.21, pp. 477–485. ISSN: 0028-1042. DOI: 10.1007/BF01507527.

— (1929). 'Crystal structure and chemical constitution'. en. In: *Transactions of the Faraday Society* 25, p. 253. ISSN: 0014-7672. DOI: 10.1039/tf9292500253.

Gordy, Walter and Thomas, W. J. Orville (1956). 'Electronegativities of the Elements'. In: *The Journal of Chemical Physics* 24.2, p. 439. ISSN: 00219606. DOI: 10.1063/1.1742493.

Grantab, Rassin, Shenoy, Vivek B and Ruoff, Rodney S (2010). 'Anomalous strength characteristics of tilt grain boundaries in graphene.' In: *Science (New York, N.Y.)* 330.6006, pp. 946–8. ISSN: 1095-9203. DOI: 10.1126/science.1196893.

Gražulis, Saulius, Chateigner, Daniel et al. (2009). 'Crystallography Open Database - an open-access collection of crystal structures.' In: *Journal of applied crystallography* 42.Pt 4, pp. 726–729. ISSN: 0021-8898. DOI: 10.1107/S0021889809016690.

Gražulis, Saulius, Daškevič, Adriana et al. (2012). 'Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration.' In: *Nucleic acids research* 40.Database issue, pp. D420–7. ISSN: 1362-4962. DOI: 10.1093/nar/gkr900.

Grimme, Stefan (2006). 'Semiempirical GGA-type density functional constructed with a long-range dispersion correction.' In: *Journal of computational chemistry* 27.15, pp. 1787–99. ISSN: 0192-8651. DOI: 10.1002/jcc.20495.

Guan, Jie, Zhu, Zhen and Tománek, David (2014a). 'High Stability of Faceted Nanotubes and Fullerenes of Multiphase Layered Phosphorus: A Computational Study'. In: *Physical Review Letters* 113.22, pp. 1–5. DOI: 10.1103/PhysRevLett.113.226801. arXiv: 1411.1141.

— (2014b). 'Tiling Phosphorene'. In: *ACS nano* 8.12, pp. 12763–12768. DOI: 10.1021/nn5059248.

Guo, Ling, Wu, Haishun and Jin, Zhihao (2004). 'First principles study of the evolution of the properties of neutral and charged phosphorus clusters'. In: *Journal of Molecular Structure: THEOCHEM* 677.1-3, pp. 59–66. ISSN: 01661280. DOI: 10.1016/j.theochem.2004.02.014.

Han, Ju-Guang and Morales, Jorge A. (2004). 'A theoretical investigation on fullerene-like phosphorus clusters'. In: *Chemical Physics Letters* 396.1-3, pp. 27–33. ISSN: 00092614. DOI: 10.1016/j.cplett.2004.07.107.

Han, Xiaoyu et al. (2014). 'Strain and orientation modulated bandgaps and effective masses of phosphorene nanoribbons'. In: *Nano Letters* 14, pp. 4607–4614. ISSN: 15306992. DOI: 10.1021/nl501658d.

Hart, Gus L. W. (2007). 'Where are nature's missing structures?' In: *Nature materials* 6.12, pp. 941–5. ISSN: 1476-1122. DOI: 10.1038/nmat2057.

— (2009). 'Verifying predictions of the L 13 crystal structure in Cd-Pt and Pd-Pt by exhaustive enumeration'. In: *Physical Review B - Condensed Matter and Materials Physics* 80.November 2008, pp. 1–5. ISSN: 10980121. DOI: 10.1103/PhysRevB.80.014106.

Hart, Gus L. W. and Forcade, Rodney W. (2008). 'Algorithm for generating derivative structures'. In: *Physical Review B - Condensed Matter and Materials Physics* 77.22, p. 224115. ISSN: 10980121. DOI: 10.1103/PhysRevB.77.224115. arXiv: 0804.3544.

Häser, Marco (1994). 'Structural Rules of Phosphorus'. In: *Journal of the American Chemical Society* 116.15, pp. 6925–6926. ISSN: 0002-7863. DOI: 10.1021/ja00094a053.

Häser, Marco, Schneider, Uwe and Ahlrichs, Reinhart (1992). 'Clusters of phosphorus: a theoretical investigation'. In: *Journal of the American Chemical Society* 114.24, pp. 9551–9559. ISSN: 0002-7863. DOI: 10.1021/ja00050a039.

Häser, Marco and Treutler, Oliver (1995). 'Calculated properties of P2, P4, and of closed-shell clusters up to P18'. In: *The Journal of Chemical Physics* 102.9, p. 3703. ISSN: 00219606. DOI: 10.1063/1.468552.

Hautier, Geoffroy, Fischer, Christopher C. et al. (2010). 'Finding Nature's Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory'. In: *Chemistry of Materials* 22.12, pp. 3762–3767. ISSN: 0897-4756. DOI: 10.1021/cm100795d.

Hautier, Geoffroy, Fischer, Chris et al. (2011). 'Data mined ionic substitutions for the discovery of new compounds.' In: *Inorganic chemistry* 50.2, pp. 656–663. ISSN: 1520-510X. DOI: 10.1021/ic102031h.

Haymet, A D J (1987). 'Theory of the Equilibrium Liquid-Solid Transition'. In: *Annual Review of Physical Chemistry* 38.1, pp. 89–108. DOI: 10.1146/annurev.pc.38.100187.000513.

Hirshfeld, F. L. (1977). 'Bonded-atom fragments for describing molecular charge densities'. In: *Theoretica Chimica Acta* 44.2, pp. 129–138. ISSN: 0040-5744. DOI: 10.1007/BF00549096.

Hoare, M. R. and McInnes, J. (1976). 'Statistical mechanics and morphology of very small atomic clusters'. en. In: *Faraday Discussions of the Chemical Society* 61, p. 12. ISSN: 0301-7249. DOI: 10.1039/dc9766100012.

Hohenberg, P. and Kohn, W. (1964). 'The Inhomogeneous Electron Gas'. In: *Phys. Rev.* 136.3B, B864. ISSN: 01631829. DOI: 10.1103/PhysRevB.7.1912.

Hopkins, Adam B., Jiao, Yang et al. (2011). 'Phase Diagram and Structural Diversity of the Densest Binary Sphere Packings'. In: *Physical Review Letters* 107.12, p. 125501. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.107.125501.

Hopkins, Adam B., Stillinger, Frank H. and Torquato, Salvatore (2012). 'Densest binary sphere packings'. In: *Physical Review E* 85.2, p. 021130. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.85.021130.

Huang, Pinshane Y et al. (2011). 'Grains and grain boundaries in single-layer graphene atomic patchwork quilts.' In: *Nature* 469.7330, pp. 389–92. ISSN: 1476-4687. DOI: 10.1038/nature09718.

Huang, Rongbin et al. (1995). 'Experimental and Theoretical Studies of Small Homoatomic Phosphorus Clusters'. In: *The Journal of Physical Chemistry* 99.5, pp. 1418–1423. ISSN: 0022-3654. DOI: 10.1021/j100005a009.

Hudson, Toby S and Harrowell, Peter (2008). 'Dense packings of hard spheres of different sizes based on filling interstices in uniform three-dimensional tilings.' EN. In: *The journal of physical chemistry. B* 112.27, pp. 8139–43. ISSN: 1520-6106. DOI: 10.1021/jp802912a.

Hultgren, Ralph et al. (1973). *Selected Values of Thermodynamic Properties of Binary Alloys, Part 2*. American Society for Metals. ISBN: 0608100714.

Hume-Rothery, William (1950). *The structure of metals and alloys*. 2nd. Institute of Metals. ISBN: 0901462381.

— (1955). *Electrons, atoms, metals and alloys*. "Metal industry," L. Cassier Co., distributor, Iliffe, p. 387.

Hunt, Neil, Jardine, Roger and Bartlett, Paul (2000). 'Superlattice formation in mixtures of hard-sphere colloids'. In: *Physical Review E* 62.1, pp. 900–913. ISSN: 1063-651X. DOI: 10.1103/PhysRevE.62.900.

Jackson, A. and Ackland, G. J. (2007). 'Lattice-switch Monte Carlo simulation for binary hard-sphere crystals'. In: *Physical Review E* 76.6, p. 066703. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.76.066703.

Jacobs, William M., Reinhardt, Aleks and Frenkel, Daan (2015). 'Rational design of self-assembly pathways for complex multicomponent structures'. In: arXiv: 1502.01351.

Jain, Anubhav et al. (2013). 'Commentary: The Materials Project: A materials genome approach to accelerating materials innovation'. In: *APL Materials* 1.1, p. 11002. ISSN: 2166532X. DOI: 10.1063/1.4812323.

Johrendt, Dirk and Pöttgen, Rainer (2008). 'Pnictide oxides: a new class of high-T(C) superconductors.' In: *Angewandte Chemie (International ed. in English)* 47.26, pp. 4782–4. ISSN: 1521-3773. DOI: 10.1002/anie.200801818.

Jones, R. O., Ganteför, G. et al. (1995). 'Structure and spectroscopy of phosphorus cluster anions: Theory (simulated annealing) and experiment (photoelectron detachment)'. In: *The Journal of Chemical Physics* 103.22, p. 9549. ISSN: 00219606. DOI: 10.1063/1.469969.

Jones, R. O. and Hohl, D. (1990). 'Structure of phosphorus clusters using simulated annealing—P2 to P8'. In: *The Journal of Chemical Physics* 92.11, p. 6710. ISSN: 00219606. DOI: 10.1063/1.458306.

Jones, R. O. and Seifert, G. (1992). 'Structure of phosphorus clusters using simulated annealing. II. P9, P10, P11, anions P2-4, P2-10, P3-11, and cations P+n to n=11'. In: *The Journal of Chemical Physics* 96.10, p. 7564. ISSN: 00219606. DOI: 10.1063/1.462408.

Karttunen, Antti J, Linnolahti, Mikko and Pakkanen, Tapani A (2008). 'Structural principles of polyhedral allotropes of phosphorus.' In: *Chemphyschem : a European journal of chemical physics and physical chemistry* 9.17, pp. 2550–8. ISSN: 1439-7641. DOI: 10.1002/cphc.200800646.

Kim, Kwanpyo et al. (2011). 'Grain boundary mapping in polycrystalline graphene.' In: *ACS nano* 5.3, pp. 2142–6. ISSN: 1936-086X. DOI: 10.1021/nn1033423.

Kirkpatrick, S, Gelatt, C D and Vecchi, M P (1983). 'Optimization by simulated annealing.' In: *Science (New York, N.Y.)* 220.4598, pp. 671–80. ISSN: 0036-8075. DOI: 10.1126/science.220.4598.671.

Klimeš, Jiří and Michaelides, Angelos (2012). *Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory*. DOI: 10.1063/1.4754130. arXiv: arXiv:1301.6960v1.

Kob, Walter and Anderson, Hans C. (1995). 'Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture I: The van Hove correlation function'. In: *Physical Review E* 51.5, pp. 4626–4641. ISSN: 1063-651X. DOI: 10.1103/PhysRevE.51.4626.

Koepke, Justin C et al. (2013). 'Atomic-scale evidence for potential barriers and strong carrier scattering at graphene grain boundaries: a scanning tunneling microscopy study.' In: *ACS nano* 7.1, pp. 75–86. ISSN: 1936-086X. DOI: 10.1021/nn302064p.

Kohn, W. and Sham, L. J. (1965). 'Self-Consistent Equations Including Exchange and Correlation Effects'. In: *Physical Review* 140.4A, A1133–A1138. ISSN: 0031-899X. DOI: 10.1103/PhysRev.140.A1133.

Kong, Chang Sun, Luo, Wei et al. (2012). 'Information-theoretic approach for the discovery of design rules for crystal chemistry'. In: *Journal of Chemical Information and Modeling* 52.7, pp. 1812–1820. ISSN: 15499596. DOI: 10.1021/ci200628z.

Kong, Chang Sun, Villars, Pierre et al. (2012). *Mapping the 'materials gene' for binary intermetallic compounds—a visualization schema for crystallographic databases*. DOI: 10.1088/1749-4699/5/1/015004.

Kou, Liangzhi et al. (2011). 'Tunable magnetism in strained graphene with topological line defect.' In: *ACS nano* 5.2, pp. 1012–7. ISSN: 1936-086X. DOI: 10.1021/nn1024175.

Kranendonk, W G T and Frenkel, Daan (1989). 'Computer simulation of solid-liquid coexistence in binary hard-sphere mixtures'. In: *Journal of Physics: Condensed Matter* 1.41, p. 7735.

— (1991a). 'Free energy calculations for solid solutions by computer simulations'. In: *Molecular Physics* 72.3, pp. 699–713. DOI: 10.1080/00268979100100511.

— (1991b). 'Thermodynamic properties of binary hard sphere mixtures'. In: *Molecular Physics* 72.3, pp. 715–733. DOI: 10.1080/00268979100100521.

Kummerfeld, Jonathan K, Hudson, Toby S and Harrowell, Peter (2008). 'The densest packing of AB binary hard-sphere homogeneous compounds across all size ratios.' EN. In: *The journal of physical chemistry. B* 112.35, pp. 10773–6. ISSN: 1520-6106. DOI: 10.1021/jp804953r.

Kuwayama, Yasuhiro et al. (2008). 'Phase relations of iron and iron-nickel alloys up to 300 GPa: Implications for composition and structure of the Earth's inner core'. In: *Earth and Planetary Science Letters* 273.3-4, pp. 379–385. ISSN: 0012821X. DOI: 10.1016/j.epsl.2008.07.001.

Laarhoven, P. J. van and Aarts, E. H. (1987). *Simulated Annealing: Theory and Applications*. Springer. ISBN: 978-9048184385.

Laio, Alessandro and Parrinello, Michele (2002). 'Escaping free-energy minima.' In: *Proceedings of the National Academy of Sciences of the United States of America* 99.20, pp. 12562–6. ISSN: 0027-8424. DOI: 10.1073/pnas.202427399.

Laves, Fritz (1955). 'Crystal structure and atomic size'. In: *Theory of Alloy Phases*, pp. 124–198.

Laves, Fritz and Witte, H (1935). 'Die Kristallstruktur des MgNi2 und seine Beziehungen zu den Typen des MgCu2 und MgZn2.' In: *Metallwirtschaft, Metallwissenschaft, Metalltechnik* 14, pp. 645–9.

Lennard-Jones, J. E. and Devonshire, A. F. (1937). 'Critical Phenomena in Gases. I'. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 163.912, pp. 53–70. ISSN: 1364-5021. DOI: 10.1098/rspa.1937.0210.

Leunissen, Mirjam E. et al. (2005). 'Ionic colloidal crystals of oppositely charged particles.' en. In: *Nature* 437.7056, pp. 235–40. ISSN: 1476-4687. DOI: 10.1038/nature03946.

Levy, M (1979). 'Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem.' In: *Proceedings of the National Academy of Sciences of the United States of America* 76.12, pp. 6062–6065. ISSN: 0027-8424. DOI: 10.1073/pnas.76.12.6062.

Li, Ting I N G et al. (2012). 'Modeling the crystallization of spherical nucleic acid nanoparticle conjugates with molecular dynamics simulations.' In: *Nano letters* 12.5, pp. 2509–14. ISSN: 1530-6992. DOI: 10.1021/nl300679e.

Li, Zheng-Lu et al. (2014). 'What are grain boundary structures in graphene?' en. In: *Nanoscale* 6.8, pp. 4309–15. ISSN: 2040-3372. DOI: 10.1039/c3nr06823d. arXiv: 1308.3964.

Lieb, Elliott H. (1983). 'Density functionals for coulomb systems'. In: *International Journal of Quantum Chemistry* 24, pp. 243–277. ISSN: 0020-7608. DOI: 10.1002/qua.560240302.

Lima-de-Faria, J et al. (1990). 'Nomenclature of inorganic structure types. Report of the International Union of Crystallography Commission on Crystallographic Nomenclature Subcommittee on the Nomenclature of Inorganic Structure Types'. In: *Acta Crystallogr., Sect. A: Found. Crystallogr.* A46, pp. 1–11. ISSN: 1600-5724. DOI: 10.1107/S0108767389008834.

Lin, Jung-Fu et al. (2002). 'Iron-silicon alloy in Earth's core?' In: *Science (New York, N.Y.)* 295.5553, pp. 313–5. ISSN: 1095-9203. DOI: 10.1126/science.1066932.

Lin, P. J. and Falicov, L. M. (1966). 'Fermi surface of arsenic'. In: *Physical Review* 142.2, pp. 441–444. ISSN: 0031899X. DOI: 10.1103/PhysRev.142.441.

Liu, Yuanyue, Xu, Fangbo et al. (2014). 'Two-Dimensional Mono-Elemental Semiconductor with Electronically Inactive Defects: The Case of Phosphorus'. In: *Nano Letters*. ISSN: 1530-6984, 1530-6992. DOI: 10.1021/nl5021393.

Liu, Yuanyue and Yakobson, Boris I (2010). 'Cones, pringles, and grain boundary landscapes in graphene topology.' In: *Nano letters* 10.6, pp. 2178–83. ISSN: 1530-6992. DOI: 10.1021/nl100988r.

Liu, Zhao-yang, Huang, Rong-bin and Zheng, Lan-sun (1996). *Bare phosphorus and binary phosphide cluster ions generated by laser ablation*. DOI: 10.1007/s004600050079.

Lord, Oliver T., Vočadlo, Lidunka et al. (2012). 'High-pressure phase transitions and equations of state in NiSi. II. Experimental results'. en. In: *Journal of Applied Crystallography* 45.4, pp. 726–737. ISSN: 0021-8898. DOI: 10.1107/S0021889812016809.

Lord, Oliver T., Wann, Elizabeth T.H. et al. (2014). 'The NiSi melting curve to 70GPa'. In: *Physics of the Earth and Planetary Interiors* 233, pp. 13–23. ISSN: 00319201. DOI: 10.1016/j.pepi.2014.05.005.

Ma, Yanming et al. (2009). 'ChemInform Abstract: Transparent Dense Sodium.' In: *ChemInform* 40.23. ISSN: 09317597. DOI: 10.1002/chin.200923005.

Macfarlane, Robert J et al. (2011). 'Nanoparticle superlattice engineering with DNA.' In: *Science (New York, N.Y.)* 334.6053, pp. 204–8. ISSN: 1095-9203. DOI: 10.1126/science.1210493.

Machlin, E., Chow, T. and Phillips, James C. (1977). 'Structural Stability of Suboctet Simple Binary Compounds'. In: *Physical Review Letters* 38.22, pp. 1292–1295. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.38.1292.

Maddox, John (1988). 'Crystals from first principles'. In: *Nature* 335.6187, pp. 201–201. ISSN: 0028-0836. DOI: 10.1038/335201a0.

Madelung, Otfried (2003). *Semiconductors: Data Handbook*. 3rd. Springer. ISBN: 978-3540404880.

Marcotte, É, Stillinger, Frank H. and Torquato, Salvatore (2013). 'Communication: Designed diamond ground state via optimized isotropic monotonic pair potentials.' In: *The Journal of chemical physics* 138.6, p. 061101. ISSN: 1089-7690. DOI: 10.1063/1.4790634.

Marqués, M. et al. (2008). 'Origin of incommensurate modulations in the high-pressure phosphorus IV phase'. In: *Physical Review B* 78.5, p. 054120. ISSN: 1098-0121. DOI: 10.1103/PhysRevB.78.054120.

Martin, Richard (2008). *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press. ISBN: 978-0521534406.

Martin, T. P. (1986). 'Compound clusters'. In: *Zeitschrift für Physik D Atoms, Molecules and Clusters* 3.2-3, pp. 211–217. ISSN: 01787683. DOI: 10.1007/BF01384809.

Martynov, A. J. and Batsanov, S. S. (1980). 'New approach to the determination of atom electronegativity'. In: *Russian Journal of Inorganic Chemistry* 25, p. 1737.

Mas-Ballesté, Rubén et al. (2011). '2D materials: to graphene and beyond.' en. In: *Nanoscale* 3.1, pp. 20–30. ISSN: 2040-3372. DOI: 10.1039/c0nr00323a.

Meel, J.A. van et al. (2008). 'Harvesting graphics power for MD simulations'. en. In: *Molecular Simulation* 34.3, pp. 259–266. ISSN: 0892-7022. DOI: 10.1080/08927020701744295.

Metropolis, Nicholas et al. (1953). 'Equation of State Calculations by Fast Computing Machines'. In: *The Journal of Chemical Physics* 21.6, p. 1087. ISSN: 00219606. DOI: 10.1063/1.1699114.

Møller, Chr. and Plesset, M. S. (1934). 'Note on an Approximation Treatment for Many-Electron Systems'. In: *Physical Review* 46.7, pp. 618–622. ISSN: 0031-899X. DOI: 10.1103/PhysRev.46.618.

Mooser, E. and Pearson, W. B. (1959). 'On the crystal chemistry of normal valence compounds'. en. In: *Acta Crystallographica* 12.12, pp. 1015–1022. ISSN: 0365110X. DOI: 10.1107/S0365110X59002857.

Morard, Guillaume et al. (2011). 'Melting of Fe–Ni–Si and Fe–Ni–S alloys at megabar pressures: implications for the core–mantle boundary temperature'. In: *Physics and Chemistry of Minerals* 38.10, pp. 767–776. ISSN: 0342-1791. DOI: 10.1007/s00269-011-0449-9.

Morris, Andrew J., Pickard, Chris J. and Needs, Richard J. (2008). 'Hydrogen/silicon complexes in silicon from computational searches'. In: *Physical Review B - Condensed Matter and Materials Physics* 78.18. ISSN: 10980121. DOI: 10.1103/PhysRevB.78.184102. arXiv: 0908.1658.

Mshumi, Chumani et al. (2014). 'Revisiting the CuPt3 prototype and the L13 structure'. In: *Acta Materialia* 73, pp. 326–336. ISSN: 13596454. DOI: 10.1016/j.actamat.2014.03.029.

Müller, Stefan and Zunger, Alex (2001). 'First-principles predictions of yet-unobserved ordered structures in the Ag-Pd phase diagram.' In: *Physical review letters* 87.16, p. 165502. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.87.165502.

Mulroue, Jack, Morris, Andrew J. and Duffy, Dorothy (2011). *Ab initio study of intrinsic defects in zirconolite*. DOI: 10.1103/PhysRevB.84.094118.

Nava, Paola and Ahlrichs, Reinhart (2008). 'Theoretical investigation of clusters of phosphorus and arsenic: fascination and temptation of high symmetries.' In: *Chemistry (Weinheim an der Bergstrasse, Germany)* 14.13, pp. 4039–45. ISSN: 0947-6539. DOI: 10.1002/chem.200701927.

Norman, N. C., ed. (1998). *Chemistry of Arsenic, Antimony and Bismuth*. Springer. ISBN: 978-0-7514-0389-3.

Novoselov, K S et al. (2004). 'Electric field effect in atomically thin carbon films.' In: *Science (New York, N.Y.)* 306.5696, pp. 666–9. ISSN: 1095-9203. DOI: 10.1126/science.1102896.

Nutt, David R. and Weller, Hilary (2009). 'Algorithm for Generating Defective Graphene Sheets'. In: *Journal of Chemical Theory and Computation* 5.7, pp. 1877–1882. ISSN: 1549-9618. DOI: 10.1021/ct900113f.

Oda, Tadao (2012). 'Convex polyhedral tilings hidden in crystals and quasicrystals'. In: *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas* 107.1, pp. 123–143. ISSN: 1578-7303. DOI: 10.1007/s13398-012-0078-2.

Oganov, Artem R., Lyakhov, Andriy O and Valle, Mario (2011). 'How evolutionary crystal structure prediction works–and why.' In: *Accounts of chemical research* 44.3, pp. 227–37. ISSN: 1520-4898. DOI: 10.1021/ar1001318.

Oganov, Artem R., Ma, Yanming et al. (2007). 'Evolutionary crystal structure prediction : overview of the USPEX method and some of its applications'. In: *Psi-k Newsletter* 84, pp. 142–171. ISSN: 1520-5215.

O'Keeffe, Michael (1999). 'Tiling by numbers'. In: *Nature* 400.August, p. 617. ISSN: 0028-0836. DOI: 10.1038/23139.

Okudera, Hiroki, Dinnebier, Robert E. and Simon, Arndt (2005). 'The crystal structure of γ -P 4 , a low temperature modification of white phosphorus'. In: *Zeitschrift für Kristallographie* 220.2-3-2005, pp. 259–264. ISSN: 0044-2968. DOI: 10.1524/zkri.220.2.259.59137.

Ong, Shyue Ping et al. (2013). 'Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis'. In: *Computational Materials Science* 68, pp. 314–319. ISSN: 09270256. DOI: 10.1016/j.commatsci.2012.10.028.

Ong, Zhun-yong et al. (2014). 'Strong Thermal Transport Anisotropy and Strain Modulation in Single-Layer Phosphorene'. In: *The Journal of Physical Chemistry C* 118, pp. 25272–25277. ISSN: 1932-7447. DOI: 10.1021/jp5079357.

Osada, Minoru and Sasaki, Takayoshi (2012). 'Two-dimensional dielectric nanosheets: novel nanoelectronics from nanocrystal building blocks.' In: *Advanced materials (Deerfield Beach, Fla.)* 24.2, pp. 210–28. ISSN: 1521-4095. DOI: 10.1002/adma.201103241.

Osters, Oliver et al. (2012). 'Synthesis and identification of metastable compounds: black arsenic–science or fiction?' In: *Angewandte Chemie (International ed. in English)* 51.12, pp. 2994–7. ISSN: 1521-3773. DOI: 10.1002/anie.201106479.

Pannetier, J. et al. (1990). 'Prediction of crystal structures from crystal chemistry rules by simulated annealing'. In: *Nature* 346.6282, pp. 343–345. ISSN: 0028-0836. DOI: 10.1038/346343a0.

Parthé, Erwin (1961). 'Space filling of crystal structures A contribution to the graphical presentation of geometrical relationships in simple crystal structures'. In: *Zeitschrift für Kristallographie* 115.1-2, pp. 52–79. ISSN: 0044-2968. DOI: 10.1524/zkri.1961.115.1-2.52.

Parthé, Erwin, Cenzual, Karin and Gladyshevskii, Roman E. (1993). 'Standardization of crystal structure data as an aid to the classification of crystal structure types'. In: *Journal of Alloys and Compounds* 197.2, pp. 291–301. ISSN: 09258388. DOI: 10.1016/0925-8388(93)90049-S.

Pauling, Linus (1960). *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*. 3rd. Cornell University Press. ISBN: 0801403332.

Pauling, Linus and Soldate, A. M. (1948). *The nature of the bonds in the iron silicide, FeSi, and related crystals*. DOI: 10.1107/S0365110X48000570.

Pearson, W. B. (1962). *Relative atomic size in semiconductor chemistry*. DOI: 10.1016/0022-3697(62)90062-8.

Pepponi, Giancarlo, Gražulis, Saulius and Chateigner, Daniel (2012). 'MPOD: A Material Property Open Database linked to structural information'. In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 284, pp. 10–14. ISSN: 0168583X. DOI: 10.1016/j.nimb.2011.08.070.

Pettifor, D. G. (1984). 'A chemical scale for crystal-structure maps'. In: *Solid State Communications* 51.1, pp. 31–34. ISSN: 00381098. DOI: 10.1016/0038-1098(84)90765-8.

— (1986). 'The structures of binary compounds. I. Phenomenological structure maps'. In: *Journal of Physics C: Solid State Physics* 19.3, pp. 285–313. ISSN: 0022-3719. DOI: 10.1088/0022-3719/19/3/002.

Pfitzner, Arno (2006). 'Phosphorus remains exciting!' In: *Angewandte Chemie (International ed. in English)* 45.5, pp. 699–700. ISSN: 1433-7851. DOI: 10.1002/anie.200503603.

Pfitzner, Arno et al. (2004). 'Phosphorus nanorods–two allotropic modifications of a long-known element.' In: *Angewandte Chemie (International ed. in English)* 43.32, pp. 4228–31. ISSN: 1433-7851. DOI: 10.1002/anie.200460244.

Phillips, James C. (1970). 'Ionicity of the Chemical Bond in Crystals'. In: *Reviews of Modern Physics* 42.3, pp. 317–356. ISSN: 0034-6861. DOI: 10.1103/RevModPhys.42.317.

Phillips, James C. and Kleinman, Leonard (1959). 'New Method for Calculating Wave Functions in Crystals and Molecules'. In: *Physical Review* 116.2, pp. 287–294. ISSN: 0031-899X. DOI: 10.1103/PhysRev.116.287.

Phillips, James C. and Van Vechten, J. (1969). 'Dielectric Classification of Crystal Structures, Ionization Potentials, and Band Structures'. In: *Physical Review Letters* 22.14, pp. 705–708. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.22.705.

Pickard, Chris J. and Needs, Richard J. (2007). *Structure of phase III of solid hydrogen*. DOI: 10.1038/nphys625.

— (2008). 'Highly compressed ammonia forms an ionic crystal.' In: *Nature materials* 7.10, pp. 775–779. ISSN: 1476-1122. DOI: 10.1038/nmat2261.

— (2011). 'Ab initio random structure searching.' In: *Journal of physics. Condensed matter : an Institute of Physics journal* 23.5, p. 053201. ISSN: 0953-8984. DOI: 10.1088/0953-8984/23/5/053201. arXiv: 1101.3987.

Portegies Zwart, Simon F., Belleman, Robert G. and Geldof, Peter M. (2007). 'High-performance direct gravitational N-body simulations on graphics processing units'. In: *New Astronomy* 12.8, pp. 641–650. ISSN: 13841076. DOI: 10.1016/j.newast.2007.05.004.

Purvis, George D. and Bartlett, Rodney J. (1982). 'A full coupled-cluster singles and doubles model: The inclusion of disconnected triples'. In: *The Journal of Chemical Physics* 76.4, pp. 1910–1918. ISSN: 0021-9606. DOI: 10.1063/1.443164.

Pusey, P. N. et al. (1989). 'Structure of crystals of hard colloidal spheres'. In: *Physical Review Letters* 63.25, pp. 2753–2756. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.63.2753.

Rabani, Eran and Egorov, S. A. (2001). 'Interactions between passivated nanoparticles in solutions: Beyond the continuum model'. In: *Journal of Chemical Physics* 115.2001, pp. 3437–3440. ISSN: 00219606. DOI: 10.1063/1.1395627.

Rechtsman, Mikael, Stillinger, Frank H. and Torquato, Salvatore (2005). 'Optimized Interactions for Targeted Self-Assembly: Application to a Honeycomb Lattice'. In: *Physical Review Letters* 95.22, p. 228301. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.95.228301.

— (2006a). 'Designed interaction potentials via inverse methods for self-assembly'. In: *Physical Review E* 73.1, p. 011406. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.73.011406.

Rechtsman, Mikael, Stillinger, Frank H. and Torquato, Salvatore (2006b). 'Self-assembly of the simple cubic lattice with an isotropic potential'. In: *Physical Review E* 74.2, p. 21404. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.74.021404.

— (2007). 'Synthetic diamond and wurtzite structures self-assemble with isotropic pair interactions'. In: *Physical Review E* 75.3, p. 031403. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.75.031403.

Redl, F X et al. (2003). 'Three-dimensional binary superlattices of magnetic nanocrystals and semiconductor quantum dots.' en. In: *Nature* 423.6943, pp. 968–71. ISSN: 1476-4687. DOI: 10.1038/nature01702.

Rodgers, David P. (1985). 'Improvements in multiprocessor system design'. In: *ACM SIGARCH Computer Architecture News* 13.3, pp. 225–231. ISSN: 01635964. DOI: 10.1145/327070.327215.

Roy, Susmit Singha and Arnold, Michael S. (2013). 'Improving Graphene Diffusion Barriers via Stacking Multiple Layers and Grain Size Engineering'. In: *Advanced Functional Materials* 23.29, pp. 3638–3644. ISSN: 1616301X. DOI: 10.1002/adfm.201203179.

Roy, Xavier et al. (2013). 'Nanoscale Atoms in Solid-State Chemistry'. In: *Science* 341.1978, pp. 157–160. ISSN: 0036-8075. DOI: 10.1126/science.1236259.

Ruck, Michael et al. (2005). 'Fibrous red phosphorus.' In: *Angewandte Chemie (International ed. in English)* 44.46, pp. 7616–9. ISSN: 1433-7851. DOI: 10.1002/anie.200503017.

Saal, James E. et al. (2013). 'Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD)'. In: *Jom* 65.11, pp. 1501–1509. ISSN: 10474838. DOI: 10.1007/s11837-013-0755-4.

Sanders, J. V. (1980). 'Close-packed structures of spheres of two different sizes I. Observations on natural opal'. In: *Philosophical Magazine A* 42.6, pp. 705–720. ISSN: 0141-8610. DOI: 10.1080/01418618008239379.

Saunders, Aaron E and Korgel, Brian A (2005). 'Observation of an AB phase in bidisperse nanocrystal superlattices.' In: *Chemphyschem : a European journal of chemical physics and physical chemistry* 6.1, pp. 61–5. ISSN: 1439-4235. DOI: 10.1002/cphc.200400382.

Schofield, A., Pusey, P. N. and Radcliffe, P. (2005). 'Stability of the binary colloidal crystals AB_{2} and AB_{13}'. In: *Physical Review E* 72.3, p. 031407. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.72.031407.

Schön, J. Christian and Jansen, Martin (1996). 'First step towards planning of syntheses in solid-state chemistry: Determination of promising structure candidates by global optimization'. In: *Angewandte Chemie International Edition in English* 35.12, pp. 1287–1304. ISSN: 1433-7851. DOI: 10.1002/anie.199612861.

— (2001a). *Determination, prediction, and understanding of structures, using the energy landscapes of chemical systems - Part I*. Article. LEKTORAT M/N, K BERBER-NERLINGER, POSTFACH 80 13 60, D-81613 MUNICH, GERMANY. DOI: 10.1524/zkri.216.6.307.20339.

— (2001b). 'Determination , prediction , and understanding of structures , using the energy landscapes of chemical systems — Part II'. In: *Zeitschrift für Kristallographie - Crystalline Materials* 216.7, pp. 361–383. ISSN: 0044-2968. DOI: 10.1524/zkri.216.7.361.20362.

Schusteritsch, Georg and Pickard, Chris J. (2014). 'Predicting interface structures: From SrTiO3 to graphene'. In: *Physical Review B - Condensed Matter and Materials Physics* 90, pp. 1–7. ISSN: 1550235X. DOI: 10.1103/PhysRevB.90.035424.

Selinger, Peter (2003). 'Potrace : a polygon-based tracing algorithm'. In: pp. 1–16.

Setyawan, Wahyu and Curtarolo, Stefano (2010). 'High-throughput electronic band structure calculations: Challenges and tools'. In: *Computational Materials Science* 49.2, pp. 299–312. ISSN: 09270256. DOI: 10.1016/j.commatsci.2010.05.010.

Shah, Parag S. et al. (2002). 'Size-Selective Dispersion of Dodecanethiol-Coated Nanocrystals in Liquid and Supercritical Ethane by Density Tuning'. In: *The Journal of Physical Chemistry B* 106.10, pp. 2545–2551. ISSN: 1520-6106. DOI: 10.1021/jp0139311.

Shao, D. F. et al. (2014). 'Electron-doped phosphorene: A potential monolayer superconductor'. In: *arXiv* 108.6, pp. 1–5. ISSN: 0295-5075. DOI: 10.1209/0295-5075/108/67004. arXiv: arXiv:1405.0092v1.

Shechtman, D. et al. (1984). 'Metallic Phase with Long-Range Orientational Order and No Translational Symmetry'. In: *Physical Review Letters* 53.20, pp. 1951–1953. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.53.1951.

Shevchenko, Elena V., Talapin, Dmitri V. et al. (2002). 'Colloidal Synthesis and Self-Assembly of CoPt 3 Nanocrystals'. In: *Journal of the American Chemical Society* 124.38, pp. 11480–11485. ISSN: 0002-7863. DOI: 10.1021/ja0259761.

Shevchenko, Elena V., Talapin, Dmitri V et al. (2006). 'Structural diversity in binary nanoparticle superlattices.' en. In: *Nature* 439.7072, pp. 55–9. ISSN: 1476-4687. DOI: 10.1038/nature04414.

Shoemaker, Clara Brink and Shoemaker, David P. (1969). 'Structural Properties of Some σ-Phase Related Phases'. In: *Developments in the Structural Chemistry of Alloy Phases*. Ed. by B C Giessen. Boston, MA: Springer US, pp. 107–139. ISBN: 978-1-4899-5566-1. DOI: 10.1007/978-1-4899-5564-7.

Shoemaker, David P., Shoemaker, Clara Brink and Wilson, F. C. (1957). 'The crystal structure of the P phase, Mo–Ni–Cr. II. Refinement of parameters and discussion of atomic coordination'. en. In: *Acta Crystallographica* 10.1, pp. 1–14. ISSN: 0365-110X. DOI: 10.1107/S0365110X57000018.

Siebert, Julien et al. (2013). 'Terrestrial accretion under oxidizing conditions.' In: *Science (New York, N.Y.)* 339.6124, pp. 1194–7. ISSN: 1095-9203. DOI: 10.1126/science.1227923.

Silas, Patricia K., Haynes, Peter D. and Yates, Jonathan R. (2013). 'Evolution of the Fermi surface of arsenic through the rhombohedral to simple-cubic phase transition: A Wannier interpolation study'. In: *Physical Review B* 88.13, p. 134103. ISSN: 1098-0121. DOI: 10.1103/PhysRevB.88.134103.

Sinha, Ashok K. (1972). 'Topologically close-packed structures of transition metal alloys'. In: *Progress in Materials Science* 15.2, pp. 81–185. ISSN: 00796425. DOI: 10.1016/0079-6425(72)90002-3.

Slater, J. (1929). 'The Theory of Complex Spectra'. In: *Physical Review* 34.10, pp. 1293–1322. ISSN: 0031-899X. DOI: 10.1103/PhysRev.34.1293.

Smith, Joseph V. (1977). 'Enumeration of 4-connected 3-dimensional nets and classification of framework silicates. I. Perpendicular linkage from simple hexagonal net'. In: *American Mineralogist* 62.7, pp. 703–709. ISSN: 0036-8075.

— (1978). 'Enumeration of 4-connected 3-dimensional nets and classification of framework silicates, II. Perpendicular and near-perpendicular linkages from 4.82, 3.122 and 4.6.12 nets.' In: *American Mineralogist* 63.1968, pp. 960–969. ISSN: 0003-004X.

— (1979). 'Enumeration of 4-connected 3-dimensional nets and classification of framework silicates, III. Combination of helix, and zigzag, crankshaft and saw chains with simple 2D net'. In: 64, pp. 551–562.

Song, Zhigong et al. (2013). 'Pseudo Hall-Petch strength reduction in polycrystalline graphene.' In: *Nano letters* 13.4, pp. 1829–33. ISSN: 1530-6992. DOI: 10.1021/nl400542n.

Srepusharawoot, Pornjuk et al. (2010). 'One-dimensional polymeric carbon structure based on five-membered rings in alkaline earth metal dicarbides BeC2 and MgC2'. In: *Physical Review B - Condensed Matter and Materials Physics* 82. ISSN: 10980121. DOI: 10.1103/PhysRevB.82.125439.

St. John, Judith and Bloch, Aaron (1974). 'Quantum-Defect Electronegativity Scale for Nontransition Elements'. In: *Physical Review Letters* 33.18, pp. 1095–1098. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.33.1095.

Stillinger, Frank H. (1999). 'Exponential multiplicity of inherent structures'. In: *Physical Review E* 59.1, pp. 48–51. ISSN: 1063-651X. DOI: 10.1103/PhysRevE.59.48.

Strong, Rachael T. (2005). 'First Principles Prediction of Crystal Structures and the EPR G-Tensor'. PhD thesis. University of Cambridge.

Tapasztó, Levente et al. (2012). 'Mapping the electronic properties of individual graphene grain boundaries'. In: *Applied Physics Letters* 100.5, p. 053114. ISSN: 00036951. DOI: 10.1063/1.3681375.

Thimm, Georg (2004). 'Crystal structures and their enumeration via quotient graphs'. In: *Zeitschrift für Kristallographie* 219.9-2004, pp. 528–536. ISSN: 0044-2968. DOI: 10.1524/zkri.219.9.528.44036.

Thomson, William (1887). 'On the division of space with minimum partitional area'. In: *Acta Mathematica* 11.151, pp. 121–134. ISSN: 00015962. DOI: 10.1007/BF02418046.

Thurn, H. and Krebs, H. (1969). 'Über Struktur und Eigenschaften der Halbmetalle. XXII. Die Kristallstruktur des Hittorfschen Phosphors'. In: *Acta Crystallographica Section B Structural*

Crystallography and Crystal Chemistry 25.1, pp. 125–135. ISSN: 05677408. DOI: 10.1107/S0567740869001853.

Tkatchenko, Alexandre and Scheffler, Matthias (2009). 'Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data'. In: *Physical Review Letters* 102.7, p. 073005. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.102.073005.

Togo, Atsushi (2015). *Spglib - a C library for finding and handling crystal symmetries.*

Tomalia, Donald A. (2009). *In quest of a systematic framework for unifying and defining nanoscience.* DOI: 10.1007/s11051-009-9632-z.

Torquato, Salvatore and Jiao, Yang (2010). 'Robust algorithm to generate a diverse class of dense disordered and ordered sphere packings via linear programming'. In: *Physical Review E* 82.6, p. 061302. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.82.061302.

Tsai, C. J. and Jordan, K. D. (1993). 'Use of an eigenmode method to locate the stationary points on the potential energy surfaces of selected argon and water clusters'. In: *The Journal of Physical Chemistry* 97.43, pp. 11227–11237. ISSN: 0022-3654. DOI: 10.1021/j100145a019.

Tsen, Adam W et al. (2012). 'Tailoring electrical transport across grain boundaries in polycrystalline graphene.' In: *Science (New York, N.Y.)* 336.6085, pp. 1143–6. ISSN: 1095-9203. DOI: 10.1126/science.1218948.

Tsuno, Kyusei, Frost, Daniel J. and Rubie, David C. (2013). 'Simultaneous partitioning of silicon and oxygen into the Earth's core during early Earth differentiation'. In: *Geophysical Research Letters* 40.1, pp. 66–71. ISSN: 00948276. DOI: 10.1029/2012GL054116.

Tuckerman, M., Berne, B. J. and Martyna, G. J. (1992). 'Reversible multiple time scale molecular dynamics'. In: *The Journal of Chemical Physics* 97.3, p. 1990. ISSN: 00219606. DOI: 10.1063/1.463137.

Turkiewicz, Ari et al. (2014). 'Assembling Hierarchical Cluster Solids with Atomic Precision'. In: *Journal of the American Chemical Society* 136.Figure 2, pp. 15873–15876. ISSN: 0002-7863. DOI: 10.1021/ja508698w.

Van Wazer, John R. (1958). *Phosphorus and Its Compounds, Volume 1.* New York: Interscience Publishers, Inc.

Vanderbilt, David (1990). 'Soft self-consistent pseudopotentials in a generalized eigenvalue formalism'. In: *Physical Review B* 41.11, pp. 7892–7895. ISSN: 01631829. DOI: 10.1103/PhysRevB.41.7892.

Velikov, Krassimir P et al. (2002). 'Layer-by-layer growth of binary colloidal crystals.' In: *Science (New York, N.Y.)* 296.5565, pp. 106–9. ISSN: 1095-9203. DOI: 10.1126/science.1067141.

Villars, Pierre (1983). 'A three-dimensional structural stability diagram for 998 binary AB intermetallic compounds'. In: *Journal of the Less Common Metals* 92.2, pp. 215–238. ISSN: 00225088. DOI: 10.1016/0022-5088(83)90489-7.

Villars, Pierre (1984a). 'A three-dimensional structural stability diagram for 1011 binary AB2 intermetallic compounds: II'. In: *Journal of the Less Common Metals* 99.1, pp. 33–43. ISSN: 00225088. DOI: 10.1016/0022-5088(84)90333-3.

— (1984b). 'Three-dimensional structural stability diagrams for 648 binary AB3 and 389 binary A3B5 intermetallic compounds: III'. In: *Journal of the Less Common Metals* 102.2, pp. 199–211. ISSN: 00225088. DOI: 10.1016/0022-5088(84)90316-3.

— (1985). 'A semiempirical approach to the prediction of compound formation for 3486 binary alloy systems'. In: *Journal of the Less Common Metals* 109.1, pp. 93–115. ISSN: 00225088. DOI: 10.1016/0022-5088(85)90110-9.

Villars, Pierre and Cenzual, Karin (2014). *Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds*. Ohio.

Vlot, By M. J. et al. (1997). 'Monte Carlo simulations of racemic liquid mixtures: thermodynamic properties and local structure'. In: *Molecular Physics* 91.1, pp. 19–30. ISSN: 0026-8976. DOI: 10.1080/002689797171706.

Vlot, Margot J. et al. (1997). 'Crystal structures of symmetric Lennard-Jones mixtures'. In: *The Journal of Chemical Physics* 107.11, p. 4345. ISSN: 00219606. DOI: 10.1063/1.474775.

Vočadlo, Lidunka (2013). 'Earth science: Core composition revealed.' In: *Nature* 495.7440, pp. 177–8. ISSN: 1476-4687. DOI: 10.1038/495177a.

Vočadlo, Lidunka, Price, Geoffrey D. and Wood, Ian G. (1999). 'Crystal structure, compressibility and possible phase transitions in epsilon-FeSi studied by first-principles pseudopotential calculations'. In: *Acta Crystallographica Section B Structural Science* 55.4, pp. 484–493. ISSN: 0108-7681. DOI: 10.1107/S0108768199001214.

Von Schnering, Hans Georg and Hoenle, Wolfgang (1988). 'Chemistry and structural chemistry of phosphides and polyphosphides. 48. Bridging chasms with polyphosphides'. In: *Chemical Reviews* 88.1, pp. 243–273. ISSN: 0009-2665. DOI: 10.1021/cr00083a012.

Wales, David J. (2004). *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*. Cambridge: Cambridge University Press. ISBN: 978-0521814157.

Wales, David J. and Doye, Jonathan P. K. (1997). 'Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms'. In: *The Journal of Physical Chemistry A* 101.28, pp. 5111–5116. ISSN: 1089-5639. DOI: 10.1021/jp970984n.

Wang, Qing Hua et al. (2012). 'Electronics and optoelectronics of two-dimensional transition metal dichalcogenides.' en. In: *Nature nanotechnology* 7.11, pp. 699–712. ISSN: 1748-3395. DOI: 10.1038/nnano.2012.193.

Warner, Jamie H et al. (2012). 'Dislocation-driven deformations in graphene.' In: *Science (New York, N.Y.)* 337.6091, pp. 209–12. ISSN: 1095-9203. DOI: 10.1126/science.1217529.

Watson, R. E. and Bennett, L. (1978). 'Transition metals: d-band hybridization, electronegativities and structural stability of intermetallic compounds'. In: *Physical Review B* 18.12, pp. 6439–6449. ISSN: 0163-1829. DOI: 10.1103/PhysRevB.18.6439.

Watson, R. E. and Bennett, L.H. H (1978). 'A Mulliken electronegativity scale and the structural stability of simple compounds'. In: *Journal of Physics and Chemistry of Solids* 39.11, pp. 1235–1242. ISSN: 00223697. DOI: 10.1016/0022-3697(78)90103-8.

Wei, Yujie et al. (2012). 'The nature of strength enhancement and weakening by pentagon-heptagon defects in graphene.' In: *Nature materials* 11.9, pp. 759–63. ISSN: 1476-1122. DOI: 10.1038/nmat3370.

Wilberg, Nils, Holleman, A. F. and Wilberg, Egon, eds. (2001). *Holleman-Wiberg's Inorganic Chemistry*. 1st. Academic Press Inc. ISBN: 978-0123526519.

Wood, Ian G. et al. (2013). 'High-pressure phase transitions and equations of state in NiSi. III. A new high-pressure phase of NiSi'. en. In: *Journal of Applied Crystallography* 46.1, pp. 14–24. ISSN: 0021-8898. DOI: 10.1107/S0021889812047085.

Xie, Jiafeng et al. (2014). 'A theoretical study of blue phosphorene nanoribbons based on first-principles calculations'. In: *Journal of Applied Physics* 116, p. 073704. ISSN: 0021-8979. DOI: 10.1063/1.4893589.

Xu, Hong and Baus, M (1992). 'A density functional study of superlattice formation in binary hard-sphere mixtures'. In: *Journal of Physics: Condensed Matter* 4.50, p. L663.

Xu, Yibin, Yamazaki, Masayoshi and Villars, Pierre (2011). 'Inorganic Materials Database for Exploring the Nature of Material'. en. In: *Japanese Journal of Applied Physics* 50.11, 11RH02. ISSN: 0021-4922. DOI: 10.1143/JJAP.50.11RH02.

Xu, Yong et al. (2013). 'Space-Charge Transfer in Hybrid Inorganic-Organic Systems'. In: *Physical Review Letters* 111.22, p. 226802. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.111.226802.

Yazyev, Oleg V and Louie, Steven G (2010). 'Electronic transport in polycrystalline graphene.' In: *Nature materials* 9.10, pp. 806–9. ISSN: 1476-1122. DOI: 10.1038/nmat2830.

Ye, Xingchen, Chen, Jun and Murray, Christopher B (2011). 'Polymorphism in self-assembled AB6 binary nanocrystal superlattices.' EN. In: *Journal of the American Chemical Society* 133.8, pp. 2613–20. ISSN: 1520-5126. DOI: 10.1021/ja108708v.

Yu, Qingkai et al. (2011). 'Control and characterization of individual grains and grain boundaries in graphene grown by chemical vapour deposition.' en. In: *Nature materials* 10.6, pp. 443–9. ISSN: 1476-1122. DOI: 10.1038/nmat3010.

Zhang, Feiwu and Oganov, Artem R. (2010). 'Iron silicides at pressures of the Earth's inner core'. In: *Geophysical Research Letters* 37.2, n/a–n/a. ISSN: 00948276. DOI: 10.1029/2009GL041224.

Zhang, Junfeng, Zhao, Jijun and Lu, Jianping (2012). 'Intrinsic strength and failure behaviors of graphene grain boundaries.' In: *ACS nano* 6.3, pp. 2704–11. ISSN: 1936-086X. DOI: 10.1021/nn3001356.

Zhang, Wen Bin et al. (2014). *Molecular nanoparticles are unique elements for macromolecular science: From "nanoatoms" to giant molecules*. DOI: 10.1021/ma401724p.

Zhang, Zhuhua et al. (2015). 'Unraveling the Sinuous Grain Boundaries in Graphene'. In: *Advanced Functional Materials* 25.3, pp. 367–373. ISSN: 1616301X. DOI: 10.1002/adfm.201403024.

Zhao, Shijun, Kang, Wei and Xue, Jianming (2014). 'The potential application of phosphorene as an anode material in Li-ion batteries'. In: *J. Mater. Chem. A* 2, pp. 19046–19052. ISSN: 2050-7488. DOI: 10.1039/C4TA04368E.

Zhu, J. H. et al. (1999). 'A thermodynamic interpretation of the size-ratio limits for laves phase formation'. In: *Metallurgical and Materials Transactions A* 30.5, pp. 1449–1452. ISSN: 1073-5623. DOI: 10.1007/s11661-999-0292-5.

Zhu, Zhen and Tománek, David (2014). 'Semiconducting Layered Blue Phosphorus: A Computational Study'. In: *Physical Review Letters* 112.17, p. 176802. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.112.176802.

Zunger, Alex (1980). 'Systematization of the stable crystal structure of all AB-type binary compounds: A pseudopotential orbital-radii approach'. In: *Physical Review B* 22.12, pp. 5839–5872. ISSN: 0163-1829. DOI: 10.1103/PhysRevB.22.5839.

Zunger, Alex and Cohen, Marvin (1978). 'Density-Functional Pseudopotential Approach to Crystal Phase Stability and Electronic Structure'. In: *Physical Review Letters* 41.1, pp. 53–56. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.41.53.

Acronyms

AIRSS *ab initio* random structure searching. 32, 33

bcc body-centered cubic. 63, 121

COD Crystallographic Open Database. 98

CVD Chemical Vapour Deposition. 83, 88

DFT density functional theory. 32, 34, 37, 41, 56, 69, 71, 76, 78, 82, 84, 85, 123

EA evolutionary algorithm. 29, 30

fcc face-centered cubic. 45, 48, 65, 67, 95, 96, 115, 117, 121

FLOP Floating Point Operation. 105, 106

FLOPS Floating Point Operations Per Second. 105

G06 Grimme. 16, 76, 78, 81, 82

GA genetic algorithm. 29, 30, 90

GB grain boundary. 83, 84, 87

GGA generalised-gradient approximation. 40, 76, 78

GPGPU General Purpose Graphics Processing Unit. 105

hcp Hexagonal Close-Packed. 32, 45, 48, 63, 65, 121

HF Hartree-Fock. 36

ICSD Inorganic Crystal Structure Database. 5, 22, 43, 44, 51, 56, 57, 64, 95, 97, 98, 100, 117, 120, 125, 129, 133

LDA local-density approximation. 40

LJ Lennard-Jones. 96, 97, 109

MNP molecular nanoparticle. 125

PES potential energy surface. 15, 23, 24, 28, 29, 31

RNG random number generator. 105, 106

RSS random structure searching. 22, 31, 32, 65

SA simulated annealing. 28, 71

TCP Tetrahedrally Close-Packed. 45, 47, 48, 50, 119

TS Tkatchenko-Scheffler. 16, 76, 78, 81, 82

UML Unified Modelling Language. 102

XC exchange-correlation. 39, 40

Index